



TUGAS AKHIR -SM141501

**PEMILIHAN JENIS ASURANSI BERDASARKAN
DEMOGRAFI CALON PEMEGANG POLIS DENGAN
METODE *NAÏVE BAYES CLASSIFIER***

Lailatul Mabadi Chaira
NRP 1212 100 036

Dosen Pembimbing
Dr. Imam Mukhlash, S.Si, MT
Drs. Nurul Hidayat, M.Kom

JURUSAN MATEMATIKA
Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Teknologi Sepuluh Nopember
Surabaya 2016



FINAL PROJECT -SM141501

***INSURANCE PRODUCT DECISION BASED ON
PROSPECTIVE CUSTOMER's DEMOGHRAPIC
PROFILES USING NAÏVE BAYES CLASSIFIER***

Lailatul Mabadi Chaira
NRP 1211 100 036

Supervisor
Dr. Imam Mukhlash, S.Si, MT
Drs. Nurul Hidayat, M.Kom

DEPARTMENT OF MATHEMATICS
Faculty of Mathematics and Natural Science
SepuluhNopember Institute of Technology
Surabaya 2016

LEMBAR PENGESAHAN

**PEMILIHAN JENIS ASURANSI BERDASARKAN
DEMOGRAFI CALON PEMEGANG POLIS
DENGAN METODE *NAÏVE BAYES CLASSIFIER***

***INSURANCE PRODUCT DECISION BASED ON
PROSPECTIVE CUSTOMER's DEMOGHRAPIC
PROFILES USING NAÏVE BAYES CLASSIFIER***

TUGAS AKHIR

Diajukan untuk memenuhi salah satu syarat
Untuk memperoleh gelar Sarjana Sains
Pada bidang studi Matematika
Program Studi S-1 Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Teknologi Sepuluh Nopember Surabaya

Oleh :
Lailatul Mabadi Chaira
NRP. 1212100036

Menyetujui,

Dosen Pembimbing II,

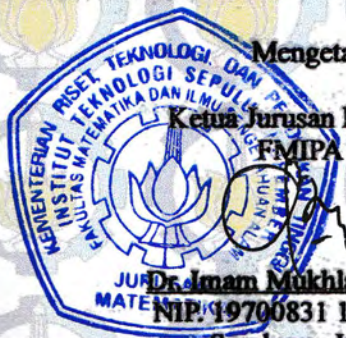
Dosen Pembimbing I,


Drs. Nurul Hidayat, M.Kom
NIP. 19630404 198903 2 001


Dr. Imam Mukhlash, S.Si, M.T
NIP. 19700831 199403 1 003

Mengetahui,

Ketua Jurusan Matematika
FMIPA ITS




Dr. Imam Mukhlash, S.Si, MT
NIP. 19700831 199403 1 003

Surabaya, Juli 2016

PEMILIHAN JENIS ASURANSI BERDASARKAN DEMOGRAFI CALON PEMEGANG POLIS DENGAN METODE *NAÏVE BAYES CLASSIFIER*

Nama Mahasiswa : Lailatul Mabadi Chaira
NRP : 1212 100 036
Jurusan : Matematika
Dosen Pembimbing 1 : Dr. Imam Mukhlash, S.Si, MT
Drs. Nurul Hidayat, M.Kom

Abstrak

Proteksi diri, baik jiwa maupun aset berharga, merupakan hal yang sangat penting untuk kehidupan individu di lingkungan yang serba beresiko saat ini. Asuransi dapat dijadikan tabungan masa depan maupun perlindungan diri ketika mengalami sakit kritis, kecelakaan, atau bahkan kematian. Perusahaan asuransi berlomba-lomba menawarkan produk asuransi yang menjanjikan. Dalam rangka bersaing dengan kompetitor lainnya dan demi memenuhi kebutuhan nasabahnya, perusahaan asuransi memerlukan strategi bisnis yang bijak dan tepat agar produknya mendapat respon positif dari calon nasabah. Oleh karena itu, diperlukan teknik *Data Mining* untuk menggali pengetahuan yang berkaitan dengan kebutuhan produk asuransi bagi calon nasabah, informasi yang dapat mempengaruhi kebutuhan asuransi seseorang adalah profil demografinya. Dalam penelitian ini digunakan Metode *Naïve Bayes Classifier* yang dinilai baik dalam hal akurasi dan efisiensi komputasi untuk membangun sebuah model klasifikasi yang akan dipakai untuk menentukan solusi produk asuransi terbaik bagi calon pemegang polis berdasarkan ciri demografinya. Hasil uji coba sistem yang dihasilkan dari 518 *record* data nasabah, Metode NBC mampu mengklasifikasi *record* dengan tingkat kinerja tertinggi sebesar 94.12%.

Kata Kunci— Asuransi, Data Mining, Klasifikasi, Naïve Bayes Classifier.

INSURANCE PRODUCT DECISION BASED ON PROSPECTIVE CUSTOMER's DEMOGHRAPIC PROFILES USING NAÏVE BAYES CLASSIFIER

Name : Lailatul Mabadi Chaira
NRP : 1212 100 036
Department : Mathematics
Supervisor : Dr. Imam Mukhlash, S.Si, MT.
Drs. Nurul Hidayat, M.Kom

Abstract

Individual protection, physicly or mentally, is very important for someone living in this risk environment. Insurance is one of individual investments and protections due to accident, critical diseases, or death. Insurance company plays a critical role in providing competitive insurance products that cover flexible features depend on customer requirements. In order to compete with other competitors and fulfill the customer needs, the company needs a wise and proper business strategy to make their offered products achieve a positive response from potential customers. Indeed, the insurance company needs a data mining technique to obtain extra knowledges on the potential customer's insurance product requirements. In this research, we proposed Naïve Bayes Classifier that assessed both in terms of accuracy and computational efficiency to develop a classification model for determining the best class solution for prospective insurance customers based on their demographic's profiles. The test result from 518 police holder records, Naïve Bayes Classifier was able to classify the data with the highest accuracy by 94.12 %.

Keywords— *Insurance, Data Mining, Classification, Naïve Bayes Classifier.*

KATA PENGANTAR

Segala Puji bagi Allah SWT yang telah memberikan karunia, rahmat dan anugerah-Nya sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul: **“Pemilihan Jenis Asuransi Berdasarkan Demografi Calon Pemegang Polis dengan Metode *Naïve Bayes Classifier*”** yang merupakan salah satu persyaratan akademis dalam menyelesaikan Program Studi S-1 pada Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Institut Teknologi Sepuluh Nopember Surabaya.

Tugas Akhir ini dapat diselesaikan dengan berkat kerjasama, bantuan, dan dukungan dari banyak pihak. Sehubungan dengan hal itu, penulis mengucapkan terima kasih kepada:

1. Dr. Imam Mukhlash, S.Si, MT selaku Ketua Jurusan Matematika ITS dan dosen pembimbing 1 Tugas Akhir ini.
2. Drs. Nurul Hidayat, M.Kom selaku dosen pembimbing 2 yang senantiasa membimbing dengan sabar dan memberikan kritik dan saran dalam penyusunan Tugas Akhir ini.
3. Drs. Komar Baihaqi, M.Si selaku Dosen Wali.
4. Dr. Budi Setiyono, S.Si, MT, Endah Rokhmah M.P, S.Si, MT, Ph.D, Dra. Farida Agustini Widjajati, MS, selaku dosen penguji Tugas Akhir ini.
5. Dr. Didik Khusnul Arif, S.Si, M.Si dan Drs. Iis Herisman, M.Sc. selaku Koordinator Tugas Akhir.
6. Seluruh jajaran dosen dan staf jurusan Matematika ITS.
7. Pihak-pihak yang membantu dalam pengerjaan Tugas Akhir ini.

Penulis menyadari bahwa Tugas Akhir ini masih jauh dari kesempurnaan. Oleh karena itu, penulis mengharapkan saran dan kritik dari pembaca. Akhir kata, semoga Tugas Akhir ini bermanfaat bagi semua pihak yang berkepentingan.

Surabaya, Juli 2016

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
LEMBAR PENGESAHAN	v
ABSTRAK	vii
ABSTRACT	ix
KATA PENGANTAR	xi
DAFTAR ISI	xiii
DAFTAR GAMBAR	xv
DAFTAR TABEL	xvii
DAFTAR LAMPIRAN	xix
 BAB I. PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	4
1.3 Batasan Masalah	4
1.4 Tujuan	4
1.5 Manfaat.....	5
1.6 Sistematika Penulisan Tugas Akhir	5
 BAB II. TINJAUAN PUSTAKA	
2.1 Penelitian Terdahulu	7
2.2 Asuransi dan Polis Asuransi	8
2.3 Data Mining	9
2.4 Klasifikasi	10
2.5 <i>Naïve Bayes Classifier</i> (NBC).....	11
2.6 <i>Knowledge Discovery in Database</i> (KDD).....	14
 BAB III. METODE PENELITIAN	
3.1 Analisis Masalah dan Studi Literatur.....	17
3.2 Pengumpulan data.....	17
3.3 Perancangan Sistem	18
3.4 Pemrosesan Awal Data	18
3.5 Pengolahan Data	18
3.6 Implementasi dan Analisa Sistem.....	21
3.7 Penarikan Kesimpulan dan Penyusunan	

Laporan Tugas Akhir	21
BAB IV. ANALISIS DAN PERANCANGAN SISTEM	
4.1 Analisis Kebutuhan Perangkat Lunak.....	23
4.2 Perancangan Sistem	23
4.2.1 Pemodelan Sistem.....	24
4.2.2 Perancangan Data.....	25
4.2.3 Perancangan Proses Kalsifikasi dengan Metode <i>Naïve Bayes Classifier</i>	28
4.2.3.1 Pemrosesan Awal Data	28
4.2.3.2 Tahap <i>Training</i>	31
4.2.3.3 Tahap Klasifikasi	37
4.3 Perancangan Antar Muka Sistem	39
4.3.1 Perancangan Struktur Menu <i>User</i>	39
4.3.2 Perancangan Antar Muka Proses Data Mining	40
BAB V. IMPLEMENTASI DAN ANALISIS SISTEM	
5.1 Implementasi Sistem.....	43
5.1.1 Implementasi Kode Program	43
5.1.2 Implementasi Data	48
5.1.3 Implementasi Antar Muka	49
5.2 Uji Coba Sistem	54
5.2.1 Lingkungan Uji Coba	55
5.2.2 Spesifikasi Dataset.....	55
5.3 Evaluasi Sistem	55
5.3.1 Jenis-jenis Pengamatan.....	55
5.3.2 Evaluasi Uji Coba.....	56
5.3.3 Analisa Hasil Evaluasi.....	64
BAB VI. KESIMPULAN DAN SARAN	
6.1 Kesimpulan	65
6.2 Saran	65
DAFTAR PUSTAKA	67
LAMPIRAN	69

DAFTAR GAMBAR

	Halaman
Gambar 2.1	Proses Klasifikasi 10
Gambar 2.2	Alur Metode <i>Naïve Bayes</i> 13
Gambar 2.3	Tahapan <i>Knowledge Discovery</i> in Database 14
Gambar 3.1	Diagram Alir Pengolahan Data Latih..... 20
Gambar 3.2	Diagram Alir Pengolahan Data Uji 20
Gambar 3.3	Diagram Alir Metodologi Penelitian..... 21
Gambar 4.1	<i>Use Case Diagram</i> Sistem Rekomendasi Asuransi..... 24
Gambar 4.2	<i>Activity Diagram</i> Sistem Rekomendasi Asuransi..... 25
Gambar 4.3	Struktur Menu <i>User</i> 40
Gambar 4.4	Perancangan Antar Muka <i>Form</i> <i>Data Mining</i> 40
Gambar 4.5	Perancangan Antar Muka Form Rekomendasi Calon Nasabah Baru..... 41
Gambar 5.1	Tampilan Sistem untuk Menampilkan Data Historis 50
Gambar 5.2	Tabel Probabilitas yang Dihasilkan dari Proses <i>Training</i> 51
Gambar 5.3	Hasil Pengujian dari Data Uji..... 52
Gambar 5.4	Form Rekomendasi untuk <i>Input</i> Data Baru..... 53
Gambar 5.5	Hasil Rekomendasi Data Baru 54
Gambar 5.6	Rekomendasi Produk SPP 56
Gambar 5.7	Rekomendasi Produk MLP 57
Gambar 5.8	Rekomendasi Produk SK 58
Gambar 5.9	Rekomendasi Produk MHP..... 59
Gambar 5.10	Grafik Prosentase Data Uji Terhadap Rata-rata Presisi 60
Gambar 5.11	Grafik Prosentase Data Uji Terhadap Rata-rata Waktu Eksekusi..... 61
Gambar 5.12	Grafik Prosentase Data Uji Terhadap Rata-rata Akurasi 63

DAFTAR TABEL

	Halaman
Tabel 4.1	Contoh <i>Dataset</i> Data Historis Nasabah
	Asuransi 29
Tabel 4.2	Frekuensi Kemunculan tiap atribut ‘Pekerjaan’. 30
Tabel 4.3	Hasil Pemrosesan Awal Data..... 31
Tabel 4.4	Tabel Peluang Atribut ‘Gender’ 32
Tabel 4.5	Tabel Peluang Atribut ‘Status Pernikahan’ 32
Tabel 4.6	Tabel Peluang Atribut ‘Status Pernikahan’ Menggunakan Teknik <i>Laplacian Smoothing</i> 33
Tabel 4.7	Tabel Peluang Atribut ‘Pekerjaan’ Menggunakan Teknik <i>Laplacian Smoothing</i> 34
Tabel 4.8	Tabel Peluang Atribut ‘Status Merokok’ Menggunakan Teknik <i>Laplacian Smoothing</i> 34
Tabel 5.1	Tabel Nasabah terhadap 10 Atribut 48
Tabel 5.2	Tabel hasil Presisi dan Running Time dengan NBC 60
Tabel 5.3	Tabel Hasil Total Akurasi Klasifikasi Dengan NBC. 62

DAFTAR LAMPIRAN

	Halaman
Lampiran A : Penurunan Formula (2.4)	69
Lampiran B : <i>Output</i> Sistem Pemilihan Jenis Asuransi	73
Lampiran C : Pengujian dengan Merubah Nilai Atribut	77
Lampiran D : Data Polis Nasabah Asuransi.....	83

BAB I

PENDAHULUAN

Pada bab ini dijelaskan mengenai latar belakang permasalahan, rumusan masalah, batasan masalah, tujuan, dan manfaat, serta sistematis penulisan dalam Tugas Akhir.

1.1 Latar Belakang Masalah

Asuransi merupakan perjanjian antar dua pihak atau lebih yang melibatkan pembayaran premi secara teratur dalam jangka tertentu sebagai ganti polis yang menjamin perlindungan apabila terjadi peristiwa yang tidak terduga seperti kecelakaan, kehilangan, kematian, kerusakan atau sakit [1]. Saat ini banyak perusahaan asuransi yang berlomba-lomba untuk menarik masyarakat agar turut serta dalam program-program asuransi yang mereka miliki. Hal ini terjadi karena kesadaran masyarakat akan masa depan mereka semakin meningkat. Dengan situasi ekonomi yang serba sulit mereka harus memikirkan bagaimana cara agar usaha mereka tetap berjalan, anak-anak mereka tetap dapat bersekolah hingga perguruan tinggi, bila mereka mengalami kecelakaan bagaimana mendapatkan biaya pengobatan dan bagaimana keluarga mereka tetap menjalani kehidupan normal meskipun sepeninggal mereka. Dengan adanya asuransi, setidaknya mengurangi biaya-biaya yang harus mereka keluarkan, selain itu mereka mendapatkan jaminan akan masa depan orang-orang yang mereka sayangi.

Penawaran asuransi tidak terbatas pada orang-orang yang terpelajar dan memiliki pendapatan besar, tetapi semua golongan masyarakat dapat turut serta didalamnya. Produk asuransi yang ditawarkan perusahaan asuransi juga semakin bermacam-macam seiring berkembangnya peradaban masyarakat dan resiko hidup yang semakin tinggi. Mulai dari produk asuransi jiwa, asuransi pendidikan, dan asuransi kesehatan.

Bermacamnya produk asuransi yang ada, seringkali membuat para agen baru perusahaan asuransi kesulitan untuk

menentukan satu produk asuransi yang sesuai dengan kebutuhan calon nasabahnya. Hal ini dikarenakan selama ini belum ada aturan-aturan baku yang digunakan untuk merekomendasikan produk asuransi bagi calon nasabah. Pemilihan jenis asuransi biasanya hanya menggunakan metode klasik, yakni perkiraan dari masing-masing agen asuransi. Hal tersebut bisa mempengaruhi kelancaran pembayaran premi di masa mendatang jika ternyata jenis asuransi yang dipilih tidak sesuai dengan kebutuhan nasabah yang bersangkutan. Alangkah baiknya pihak perusahaan asuransi dapat membantu menentukan jenis asuransi yang sesuai untuk calon pemegang polis, sehingga tidak perlu menghabiskan waktu hanya untuk menganalisis semua jenis asuransi yang ada sampai menemukan yang sesuai dengan kebutuhan calon nasabah. Hal ini melahirkan suatu kebutuhan terhadap teknologi yang dapat memanfaatkannya dalam menggali pengetahuan-pengetahuan baru, yang dapat membantu dalam penerapan strategi bisnis asuransi. Dengan memanfaatkan jumlah data yang sangat besar, pihak perusahaan tentunya dapat menemukan beragam informasi. Salah satu informasi yang dapat dihasilkan yaitu informasi mengenai ciri demografi nasabah terhadap produk asuransi yang dipilihnya. Informasi yang dihasilkan sangat penting bagi suatu perusahaan asuransi, dimana dengan adanya informasi kriteria nasabah perusahaan asuransi dapat mengambil keputusan dalam menerapkan strategi yang tepat untuk menawarkan produk kepada calon nasabah berdasarkan klasifikasi demografi calon nasabah dan pola kecenderungan nasabah terdahulu.

Klasifikasi adalah proses menemukan kumpulan pola atau fungsi-fungsi yang mendeskripsikan dan memisahkan kelas data satu dengan lainnya yang dapat digunakan untuk memprediksi data yang belum memiliki kelas data tertentu [5]. Dalam beberapa tahun terakhir, sudah ada beberapa penelitian untuk klasifikasi penentuan jenis asuransi. Salah satunya pernah dilakukan oleh Azuraliza Abu Bakar, dkk (2010) dengan menggunakan pendekatan klasifikasi *Associative Rule* untuk menentukan produk asuransi yang sesuai bagi calon nasabah asuransi berdasarkan data diri calon nasabah,

dimana keakurasian yang dicapai dengan metode ini hanya berkisar 58% [2]. Selanjutnya penelitian dari Bi Xing, Li Jian, dan Huang Feng-wen (2009) mengklasifikasikan nasabah asuransi menggunakan algoritma *Backpropagation*, penelitian ini menunjukkan kinerja sistem yang hampir sama antara *input* dan *output*-nya. Namun, sistem ini belum memberikan rekomendasi produk lain yang kemungkinan juga cocok untuk calon nasabah [3].

Selain menggunakan pendekatan *Associative Rule* dan Algoritma *Backpropagation*, teknik klasifikasi dapat juga menggunakan Metode *Naïve Bayes Classifier* (NBC). Metode klasifikasi ini diturunkan dari penerapan teorema Bayes dengan asumsi saling bebas satu sama lain berdasarkan atribut kelas [6]. Pada penelitian sebelumnya, metode *Naïve Bayes* pernah diimplementasikan oleh Bustami (2013) untuk membuat sebuah sistem penentu kelancaran pembayaran premi asuransi. Dalam penelitian ini, performa *Naïve Bayes* memiliki keakurasian yang cukup baik dalam mengklasifikasikan data nasabah [8]. Dalam jurnal lain yang ditulis oleh Sri Kusumadewi (2009) *Naïve Bayes Classifier* diaplikasikan untuk penentuan status gizi seseorang berdasarkan hasil pengukuran antropometri. Performa yang dihasilkan dari *Naïve Bayes Classifier* pada penelitian ini memiliki kinerja sistem sekitar 93,2% [6]. Selain itu, Murat karatabak (2015) juga pernah menggunakan metode *Naïve Bayes Classifier* dalam mendeteksi penyakit kanker payudara. Pada penelitiannya, Murat memaparkan hasil yang cukup signifikan bahwa metode NBC secara tepat mendeteksi kanker payudara dengan keakuratan sebesar 96,17% [4]. Dari beberapa hasil penelitian yang pernah dilakukan, penulis menyimpulkan bahwa Metode *Naïve Bayes Classifier* (NBC) dinilai berpotensi baik dalam mengklasifikasi data dibandingkan beberapa metode pengklasifikasian yang lain dalam hal akurasi dan efisiensi komputasi dikarenakan sifat keindependensian fitur-fiturnya.

Berdasarkan latar belakang tersebut, penulis terinspirasi untuk mengatasi masalah tersebut dengan mengembangkan sistem

“Pemilihan Jenis Asuransi Berdasarkan Demografi Calon Pemegang Polis Menggunakan Metode *Naïve Bayes Classifier*”.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang tersebut, permasalahan yang dirumuskan dalam Tugas Akhir ini adalah sebagai berikut:

1. Bagaimana mengimplementasikan metode *Naïve Bayes Classifier* untuk pemilihan jenis asuransi?
2. Bagaimana kinerja metode *Naïve Bayes Classifier* dalam mengklasifikasikan data nasabah asuransi?

1.3 Batasan Masalah

Pembahasan pada Tugas Akhir ini dibatasi pada beberapa hal berikut :

1. Data yang digunakan merupakan profil demografi nasabah aktif dari kantor AXA Financial Indonesia cabang Gresik sejak bulan Januari 2014 - Januari 2016 (2 tahun).
2. Data demografi nasabah yang akan dijadikan atribut adalah produk asuransi, jenis kelamin, usia, pekerjaan, bidang pekerjaan, pendapatan, status pernikahan, jumlah anak, status merokok, dan ada tidaknya riwayat kesehatan *Pre-Existing*.
3. Jenis asuransi yang digunakan adalah Smart Kidz (SK), Maestrolink Plus (MLP), Super Protection Plan (SPP), dan Maestro Hospital Plan (MHP).
4. Perangkat lunak yang dibuat Tidak menyangkut perhitungan premi dan uang pertanggungan untuk masing-masing jenis asuransi yang direkomendasikan.

1.4 Tujuan

Tujuan umum dari Tugas Akhir ini adalah mengimplementasikan metode *Naïve Bayes Classifier* untuk pemilihan jenis asuransi berdasarkan demografi calon pemegang polis.

Tujuan khusus dari Tugas Akhir ini adalah sebagai berikut:

1. Menghasilkan suatu sistem rekomendasi jenis asuransi berdasarkan demografi calon pemegang polis dengan Metode *Naïve Bayes Classifier*
2. Mengetahui kinerja Metode *Naïve Bayes Classifier* dalam mengklasifikasi data nasabah asuransi.

1.5 Manfaat

Manfaat dalam Tugas Akhir ini adalah:

1. Bagi Perusahaan
Dengan adanya sistem ini, diharapkan dapat membantu menyediakan pengetahuan dan informasi yang mendukung untuk merekomendasikan jenis asuransi yang sesuai dengan ciri demografi calon nasabah berdasarkan data nasabah terdahulu.
2. Bagi Peneliti dan pembaca
Dapat memberikan wawasan dan pengembangan serta penerapan data mining dengan menggunakan Metode *Naïve Bayes Classifier* dalam bidang asuransi.

1.6 Sistematika Penulisan Tugas Akhir

Sistematika penulisan Tugas Akhir ini adalah:

BAB I Pendahuluan

Bab ini berisi diskripsi umum isi Tugas Akhir yang meliputi latar belakang masalah, batasan masalah, tujuan, manfaat, dan sistematika penulisan Tugas Akhir.

BAB II Tinjauan Pustaka

Pada bab ini dibahas dasar teori yang menjadi landasan dalam penyusunan Tugas Akhir. Dasar teori ini terdiri dari penjelasan tentang asuransi, *Data Mining*, Klasifikasi, dan teori tentang *Naïve Bayes Classifier*.

BAB III Metode Penelitian

Pada bab ini dijelaskan metode penyelesaian permasalahan yang telah dirumuskan pada Bab I.

BAB IV Analisis dan Perancangan Sistem

Bab ini berisi pembahasan mengenai kebutuhan perangkat lunak dan perancangan sistem yang meliputi pemodelan sistem, perancangan data, perancangan proses klasifikasi dengan *Naïve Bayes Classifier*, dan perancangan antar muka.

BAB V Implementasi dan Analisa Sistem

Pada bab ini akan dibahas mengenai pembuatan perangkat lunak, uji coba perangkat lunak dan analisis terhadap hasil uji coba yang dilakukan.

BAB VI Penutup

Bab ini berisi kesimpulan dan saran.

BAB II TINJAUAN PUSTAKA

Pada bab ini dibahas teori-teori yang terkait dengan permasalahan dalam Tugas Akhir ini. Pertama, membahas mengenai penelitian terdahulu, asuransi, *data mining*, pengertian klasifikasi, *Naïve Bayes Classifier*, serta membahas *Knowledge Discovery in Database* (KDD).

2.1 Penelitian Terdahulu

Penelitian terkait dengan penentuan jenis asuransi bagi calon nasabah asuransi diantaranya adalah penelitian yang dilakukan oleh Azuralia Abu bakar, dkk (2010) dengan judul *Development of Knowledge Model for Insurance Product Decision using the Associative Classification Approach* [2]. Penelitian ini bertujuan untuk memberikan informasi pemilihan jenis asuransi yang sesuai bagi nasabah dengan membangkitkan aturan-aturan yang terbentuk dari Pendekatan *Associative Rule*. Penelitian ini menggunakan 200.000 *record* nasabah dari suatu perusahaan asuransi di Malaysia, dengan memberikan nilai *threshold* awal sebesar 0.01 untuk *minimum support* dan 0.8 untuk *minimum confidence*. Akan tetapi, metode ini hanya berhasil mengklasifikasikan data secara benar dengan akurasi sebesar 58%.

Bi Xing, Li Jian, dan Huang Feng-wen (2009) dalam makalah yang berjudul *The Application of Improved BP Algorithm in Costumer Classification of Life Insurance* [3]. Sama halnya dengan penelitian Azuralia Abu bakar dkk, penelitian ini juga bertujuan untuk merekomendasikan jenis asuransi yang sesuai bagi calon nasabah dengan mengembangkan Algoritma *Backpropagation*. Penelitian ini menggunakan 200 data pelatihan, dengan inisialisasi *input node* sebanyak $m=5$, *hidden node* sebanyak $n=8$, dan *output node* diberikan $k=1$, serta untuk koefisien *interrelated* $C_1=0.75$ dan derajat radiasi sebesar $C_2=0.0005$, maka *output* yang dihasilkan untuk data pengujian akan mendekati dengan *input* data yang bersangkutan. Sehingga dapat dikatakan hasilnya adalah sistem konsultasi yang kredibel bagi

nasabah asuransi. Namun demikian, sistem ini tidak menyajikan urutan produk yang patut direkomendasikan.

Sementara itu, ada beberapa penelitian terkait pengimplementasian Metode *Naïve Bayes Classifier* diantaranya yakni penelitian yang dilakukan Bustami (2013) dengan judul Penerapan Algoritma *Naïve Bayes* Untuk Mengklasifikasi Data Nasabah Asuransi [8]. Sistem klasifikasi data nasabah ini digunakan untuk menampilkan informasi klasifikasi kelancaran calon nasabah dalam membayar premi asuransi dengan menggunakan algoritma *Naïve Bayes*. Hasilnya dari 20 *dataset*, diperoleh tabel probabilitas masing-masing kelas yang nanti digunakan untuk menghitung probabilitas posterior maksimum sebagai penentuan kelas untuk data baru.

Sri Kusumadewi (2009) dalam makalah yang berjudul *Klasifikasi Status Gizi Menggunakan Naïve Bayes Classification* [6], menyatakan bahwa algoritma NBC dapat digunakan sebagai salah satu metode untuk klasifikasi status gizi berdasarkan hasil pengukuran antropometri. Ada 47 *dataset* yang akan diklasifikasi ke dalam 4 kelompok status gizi yakni kurang, normal, obesitas sedang, dan obesitas berat. Dari hasil uji coba, model sistem yang dibangun memiliki kinerja yang baik karena hasil pengujian menunjukkan total kinerja sebesar 0,932 atau 93,2%.

Murat Karabatak (2015) dalam makalah yang berjudul *A New Classifier for Breast Cancer detection based on Naïve Bayesian* [4], Murat melakukan penelitian dengan menerapkan algoritma NBC untuk mengklasifikasikan pasien yang terdiagnosa kanker payudara ke dalam 2 stadium, yaitu tergolong kanker jinak atau kanker ganas. Dari 65% *dataset* kanker ganas dan 35% *dataset* kanker jinak, hasil penelitiannya menunjukkan bahwa keakurasian klasifikasi dengan menggunakan metode NBC yang didapat mencapai 96,17%.

2.2 Asuransi dan Polis Asuransi

Definisi asuransi merupakan suatu alat untuk mengurangi risiko yang melekat pada perekonomian, dengan cara menggabungkan sejumlah unit-unit yang terkena risiko yang sama

atau hampir sama, dalam jumlah yang cukup besar, agar probabilitas kerugiannya dapat diramalkan dan bila kerugian yang diramalkan terjadi akan dibagi secara proposional oleh semua pihak dalam gabungan itu [1].

Polis asuransi adalah sebuah perjanjian yang menjamin pembayaran sejumlah dana atas kematian pihak bertanggung atau keadaan lain yang telah disebutkan dalam kontrak perjanjian. Secara umum polis merupakan bukti tertulis untuk perjanjian asuransi, dengan ketentuan: dibuat dengan maksud baik dari kedua belah pihak yang mengadakan perjanjian, dituliskan/disebutkan dengan tegas dan jelas mengenai hal-hal yang diperjanjikan oleh kedua belah pihak, hak masing-masing pihak, sanksi atas pelanggaran perjanjian, dan sebagainya, serta redaksinya harus disusun sedemikian rupa sehingga dengan mudah dapat ditangkap maksud dari perjanjian itu, juga tidak memberi peluang untuk menyalahafsirkannya [1].

2.3 Data Mining

Data Mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. *Data Mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan atau *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar [6].

Kemajuan luar biasa yang terus berlanjut dalam bidang *data mining* didorong oleh beberapa faktor, antara lain, pertumbuhan yang cepat dalam kumpulan data, penyimpanan data dalam *data warehouse* sehingga seluruh perusahaan memiliki akses ke dalam *database* yang andal, adanya peningkatan akses data melalui navigasi *web* dan intranet, tekanan kompetisi bisnis untuk meningkatkan penguasaan pasar dalam globalisasi ekonomi, perkembangan teknologi perangkat lunak untuk *data mining*, perkembangan yang hebat dalam kemampuan komputasi dan pengembangan kapasitas media penyimpanan [5].

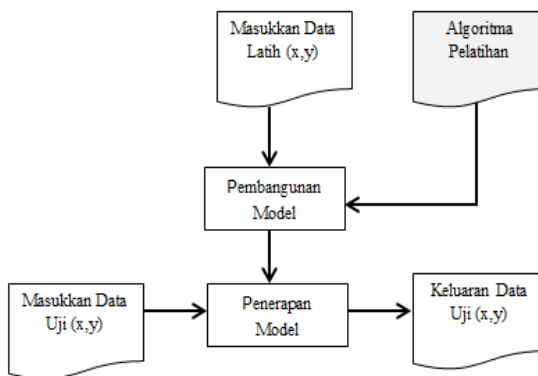
Dalam prosesnya *data mining* menggunakan apa yang dihasilkan oleh *data warehouse*. *Data warehouse* adalah kumpulan data dari

berbagai sumber internal maupun eksternal yang dikumpulkan menjadi satu dalam tempat penyimpanan yang berukuran besar yang bisa digunakan untuk pengambilan keputusan, dan bersifat berorientasi subjek, terintegrasi, *time-variant*, dan *non-volatile*. *Data warehouse* bertugas untuk menarik data dari basis data mentah untuk memberikan hasil data yang nantinya digunakan untuk proses *data mining*.

2.4 Klasifikasi

Klasifikasi adalah suatu fungsionalitas *data mining* yang akan menghasilkan model untuk memprediksi kelas atau kategori dari objek-objek di dalam basis data. Klasifikasi merupakan proses yang terdiri dari dua tahap, yaitu tahap pembelajaran dan tahap pengklasifikasian.

Pada tahap pelatihan, sebuah algoritma klasifikasi akan membangun sebuah model klasifikasi dengan cara menganalisis *data training*. Tahap pembelajaran dapat juga dipandang sebagai tahap pembentukan fungsi atau pemetaan $Y = F(X)$ dimana Y adalah kelas hasil prediksi dan X adalah tuple yang ingin diprediksi kelasnya. Selanjutnya pada tahap pengklasifikasian, model yang telah dihasilkan akan digunakan untuk melakukan klasifikasi. Model itu sendiri bisa berupa aturan “*jika-maka*”, berupa pohon keputusan, atau formula matematis [6].



Gambar 2.1 Proses Klasifikasi

2.5 *Naïve Bayes Classifier* (NBC)

Metode *Naïve Bayes* merupakan suatu metode klasifikasi yang menggunakan perhitungan probabilitas. Sedangkan Klasifikasi Bayes adalah klasifikasi statistik yang dapat memprediksi kelas suatu anggota probabilitas. Konsep dasar yang digunakan pada Metode *Naïve Bayes* adalah Teorema Bayes yang dinyatakan pertama kali oleh Thomas Bayes [6]. Nilai probabilitas yang digunakan dinyatakan secara sederhana sebagai berikut:

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)} \dots\dots\dots (2.1)$$

Dimana,

$P(X|Y)$ = Probabilitas akhir bersyarat (*conditional probability*) suatu hipotesis X terjadi jika diberikan bukti (*evidence*) Y terjadi

$P(Y|X)$ = Probabilitas sebuah bukti Y terjadi akan mempengaruhi hipotesis X

$P(X)$ = Probabilitas awal (priori) hipotesis X tanpa memandang bukti apapun

$P(Y)$ = Probabilitas awal (priori) bukti Y terjadi tanpa memandang hipotesis/bukti yang lain

Untuk Klasifikasi Bayes sederhana yang lebih dikenal sebagai *Naïve Bayesian Classifier* (NBC) dapat diasumsikan bahwa efek dari suatu nilai atribut sebuah kelas yang diberikan adalah bebas dari atribut-atribut lain. Ciri utama dari *Naïve Bayes Classifier* ini adalah asumsi yang sangat kuat (naif) akan independensi dari masing-masing kondisi/kejadian, dimana diasumsikan bahwa setiap atribut prediktor bersifat saling lepas satu sama lain berdasarkan atribut kelas. Kaitan antara *Naïve Bayes* dengan klasifikasi, korelasi hipotesis dan bukti klasifikasi adalah bahwa hipotesis dalam teorema Bayes merupakan label kelas yang menjadi target pemetaan dalam klasifikasi, sedangkan bukti merupakan fitur-fitur yang menjadikan masukan dalam model klasifikasi. Jika X adalah vektor masukan yang berisi fitur, maka Y adalah label kelas, *Naïve Bayes* dituliskan dengan $P(X|Y)$. Notasi

tersebut berarti probabilitas label kelas Y didapat setelah fitur-fitur X diamati. Notasi ini disebut probabilitas akhir (*Posterior probability*) untuk Y , sedangkan $P(Y)$ disebut probabilitas awal (*Prior Probability*) Y .

Dalam proses pelatihan data, dilakukan perhitungan probabilitas kondisional $P(X|Y)$ untuk setiap kombinasi X dan Y berdasarkan informasi yang didapat dari data latih. Dengan membangun model tersebut, suatu data uji X' dapat diklasifikasikan dengan mencari nilai Y' dan memaksimalkan $P(Y'|X')$ yang didapat.

Formula *Naïve Bayes Classifier* adalah sebagai berikut:

$$P(Y|X) = \frac{P(Y) \prod_{i=1}^q P(X_i|Y)}{P(X)} \dots\dots\dots (2.2)$$

$P(Y|X)$ adalah probabilitas data dengan vektor X pada kelas Y . $P(Y)$ adalah probabilitas awal kelas Y . $\prod_{i=1}^q P(X_i|Y)$ merupakan probabilitas independen kelas Y dari semua fitur dalam vektor X . Nilai $P(X)$ selalu tetap pada semua kelas Y untuk satu sampel data, sehingga dalam perhitungan prediksi tinggal dihitung bagian $P(Y) \prod_{i=1}^q P(X_i|Y)$ dengan memilih nilai yang terbesar menjadi kelas yang dipilih sebagai hasil prediksi. Untuk menentukan pilihan kelas, digunakan peluang posterior maksimal dari seluruh y dalam Y dengan fungsi:

$$\underset{y \in Y}{\operatorname{argmax}} P(y) \prod_{i=1}^q P(X_i|y) \dots\dots\dots (2.3)$$

Umumnya, Metode *Bayes* mudah dihitung untuk fitur bertipe kategorik. Namun untuk fitur dengan tipe numerik (kontinyu) ada perlakuan khusus sebelum dimasukkan dalam *Naïve Bayes*. Ada 2 cara yang dapat dilakukan, yaitu:

1. Melakukan diskretisasi pada setiap fitur kontinu dan mengganti nilai fitur kontinu tersebut dengan nilai interval diskret. Pendekatan ini dilakukan dengan mentransformasikan fitur kontinu ke dalam fitur ordinal.
2. Mengasumsikan bentuk tertentu dari distribusi probabilitas untuk fitur kontinu dan memperkirakan parameter distribusi dengan data pelatihan. Distribusi *Gaussian* biasanya dipilih

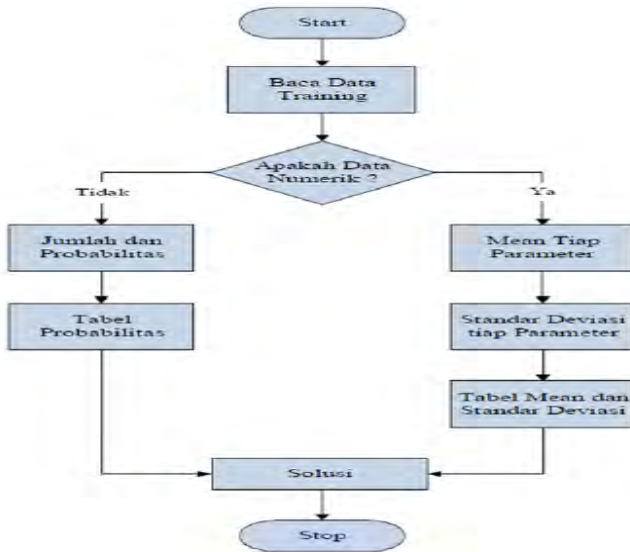
untuk merepresentasikan probabilitas bersyarat dari fitur numerik/kontinyu pada sebuah kelas $P(X_i|Y)$, sedangkan Distribusi *Gaussian* dikarakteristikkan dengan dua parameter, yakni *mean* (μ) dan *varian* (σ^2). Untuk setiap kelas y_j , probabilitas bersyarat kelas y_j untuk fitur x_i adalah:

$$P(X_i = x_i|Y = y_j) = \frac{1}{\sigma_{ij}\sqrt{2\pi}} e^{-\frac{(x_{ij}-\mu_{ij})^2}{2\sigma_{ij}^2}} \dots\dots (2.4)$$

Parameter μ_{ij} bisa didapat dari *mean* sampel $X_i(\bar{x})$ dari semua data latih yang menjadi milik kelas y_j , sedangkan σ_{ij}^2 dapat diperkirakan dari *varian* sampel (s^2) dari data latih yang menjadi milik kelas y_j . Dimana,

$$\mu_{ij} = \frac{\sum_{i=1}^n x_{ij}}{n_j} \text{ dan } \sigma_{ij}^2 = \frac{\sum_{i=1}^n (x_{ij}-\mu_{ij})^2}{n_j} \dots\dots\dots(2.5)$$

Lebih jelasnya, Alur dari metode *Naïve Bayes* dapat dilihat pada gambar 2.2 sebagai berikut [6]:

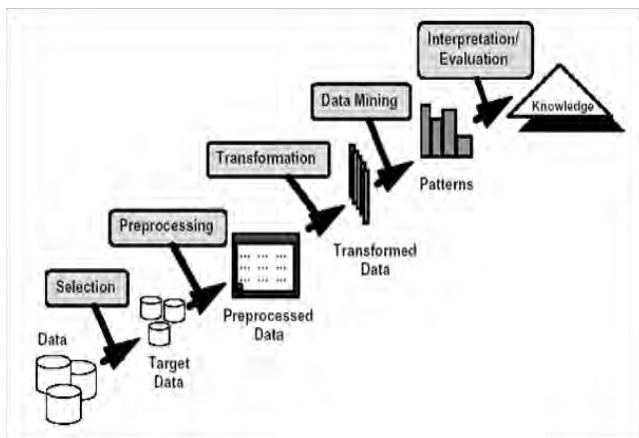


Gambar 2.2 Alur Metode *Naïve Bayes*

2.6 Knowledge Discovery in Databases (KDD)

Proses KDD adalah proses menggunakan metode *data mining* untuk mengekstrak pengetahuan apa yang dianggap sesuai dengan spesifikasi ukuran dan batas, menggunakan *database* bersama dengan *preprocessing* yang diperlukan, pengambilan sampel dan transformasi dari *database* [7].

Istilah *data mining* dan *Knowledge discovery in database* (KDD) seringkali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Salah satu tahapan dalam keseluruhan proses KDD adalah *data mining*. Proses KDD secara garis besar dapat dijelaskan sebagai berikut [7].



Gambar 2.3 Tahapan *Knowledge Discovery in Database*

Gambar 2.3 menunjukkan proses tahapan KDD dimana keterangan setiap tahapnya dijelaskan sebagai berikut:

1. *Data Selection*, pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses *data mining*, disimpan dalam suatu berkas, terpisah dari basis data operasional

2. *Preprocessing/Cleaning*, sebelum proses *data mining* dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus KDD. Proses *cleaning* mencakup pembuangan duplikasi data, pemeriksaan data yang inkonsisten, dan perbaikan kesalahan pada data, seperti kesalahan cetak.
3. *Transformation, coding* adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining*. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada sejenis atau pola informasi yang akan dicari dalam basis data.
4. *Data mining*, merupakan proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.
5. *Interpretation/Evaluation*, pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

BAB III

METODE PENELITIAN

Pada bab ini dijelaskan metode yang digunakan dalam Tugas Akhir agar proses pengerjaan dapat terstruktur dengan baik dan dapat mencapai tujuan yang telah ditetapkan sebelumnya. Tahapan dari metode penelitian ini digambarkan secara terperinci dengan diagram alur seperti pada Gambar 3.3.

3.1 Analisis Masalah dan Studi Literatur

Tahap ini adalah langkah awal untuk menentukan rumusan masalah dari penelitian. Dalam hal ini mengamati permasalahan yang berhubungan dengan atribut-atribut yang mempengaruhi pemilihan jenis asuransi. Dari permasalahan-permasalahan yang ada, selanjutnya dianalisa untuk mengetahui bagaimana cara penyelesaian terhadap masalah tersebut dan menentukan ruang lingkup permasalahan yang akan diteliti. Mempelajari dasar teori dari berbagai literatur mengenai penerapan Metode *Naïve Bayes Classifier*, konsep dan teori *data mining* serta literatur terkait asuransi, melalui jurnal-jurnal, buku pustaka serta wawancara langsung ke pihak terkait dengan perusahaan asuransi agar mendapatkan dasar pengetahuan untuk melakukan penelitian selanjutnya.

Dari permasalahan dan tujuan yang telah dirumuskan selanjutnya dilakukan studi literatur untuk mendukung pengerjaan Tugas Akhir. Studi literatur dilakukan terhadap jurnal-jurnal ilmiah, tugas akhir, dan buku-buku penunjang, serta referensi dari internet yang terkait dengan *Naïve Bayes Classifier*, seperti yang telah tercantum dalam daftar pustaka.

3.2 Pengumpulan Data

Pengumpulan data dilakukan untuk mendapatkan data yang dibutuhkan dalam pengerjaan Tugas Akhir. Data yang digunakan yaitu data demografi nasabah salah satu asuransi yang berada di Indonesia dan memiliki kantor cabang di Gresik. Data diperoleh

dengan melakukan wawancara karyawan dan mengirimkan surat pengantar dari jurusan Matematika ITS kepada SOE (*Secretary Office Executive*) perusahaan asuransi yang dituju, kemudian pihak asuransi memberikan persetujuan tentang detail data profil demografi nasabah yang boleh diambil sebagai bahan penelitian. Dalam pengerjaan Tugas Akhir ini, data demografi nasabah yang dijadikan sebagai atribut antara lain produk asuransi, jenis kelamin, usia, pekerjaan, bidang pekerjaan, pendapatan, status pernikahan, jumlah anak, status merokok, dan riwayat kesehatan.

3.3 Perancangan Sistem

Tujuan dari perancangan sistem adalah untuk memenuhi kebutuhan *user* mengenai gambaran yang jelas tentang perancangan sistem yang akan dibuat serta diimplementasikan. Untuk memulai membangun suatu program berupa aplikasi untuk pemilihan jenis asuransi yang sesuai dengan kriteria demografi calon nasabah, maka penulis terlebih dahulu merencanakan alur kerja berdasarkan kebutuhan *user* yang menggunakan aplikasi ini.

3.4 Pemrosesan Awal Data

Proses awal setelah pengumpulan data nasabah asuransi. Pada proses ini data nasabah asuransi akan dinormalisasi, dimulai dengan proses *cleaning data*, dimana penduplikasian data dihilangkan dan penanganan nilai atribut yang hilang. Tujuan dari proses ini adalah untuk mereduksi *noise* (gangguan) data yang hendak di-*mining*. Kemudian, akan dilakukan transformasi data jika dibutuhkan, yaitu dengan merubah tipe suatu atribut sehingga sesuai dengan proses *data mining* dan mengurangi jumlah atribut yang tidak diperlukan dalam proses data mining yang akan dilakukan.

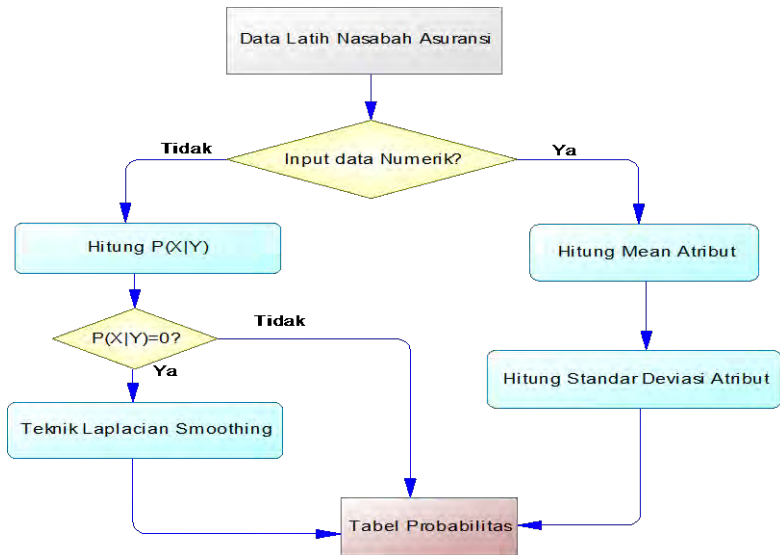
3.5 Pengolahan Data

Pada tahap ini akan dilakukan pengolahan data untuk mendapatkan hasil klasifikasi nasabah terhadap jenis asuransi.

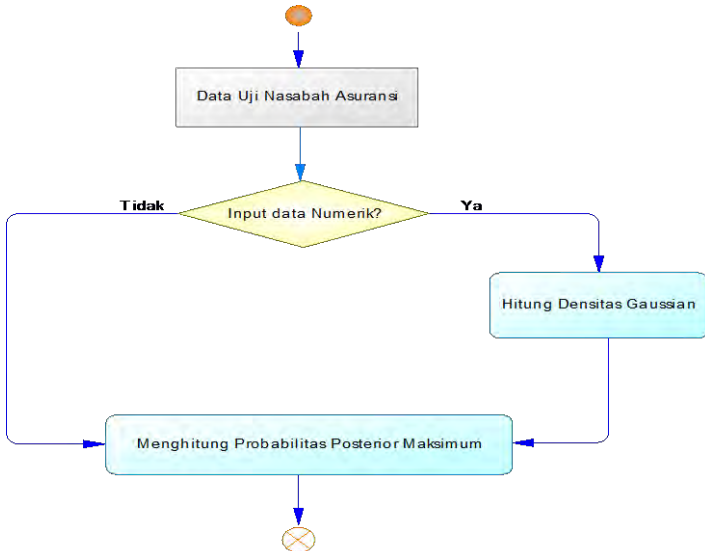
Langkah-langkah pengklasifikasian menggunakan Metode *Naïve Bayes Classifier* adalah sebagai berikut:

1. Membagi proporsi data nasabah asuransi menjadi 2 kelompok data. Yaitu 90% sebagai data latih dan 10% sebagai data uji.
 - a. Proses *Training*
 1. Membaca data nasabah asuransi yang akan digunakan sebagai data latih.
 2. Memeriksa bentuk dari setiap kriteria nasabah asuransi, baik yang bertipe numerik/kontinyu maupun yang berbentuk kategorik. Data numerik merupakan data yang memiliki *input* berupa angka atau nominal. Sedangkan data kategorik merupakan data yang sudah memiliki klasifikasi nilai tertentu.
 3. Jika data berbentuk numerik, maka dihitung *mean* dan standar deviasi tiap atribut. Jika data bersifat kategorik, maka langkah selanjutnya yaitu menghitung probabilitas *prior* tiap kelasnya, dan akhirnya diperoleh tabel probabilitas.
 - b. Proses *Testing*
 1. Mengevaluasi data uji dengan menghitung probabilitas bersyarat setiap kriteria.
 2. Menghitung probabilitas posterior pada data uji.
 3. Menentukan probabilitas posterior maksimum sebagai solusi kelas klasifikasi data uji. Dimana data uji akan diklasifikasikan ke dalam salah satu dari 4 (empat) jenis asuransi yakni Smart Kidz (SK), Maestrolink Plus (MLP), Super Protection Plan (SPP) dan Maestro Hospital Plan (MHP).
2. Melakukan analisa keakurasian klasifikasi menggunakan Metode *Naïve Bayes Classifier* dan penarikan kesimpulan.

Secara rinci, skema pengolahan data latih dan data uji akan digambarkan pada Gambar 3.1 dan Gambar 3.2:



Gambar 3.1 Diagram Alir Pengolahan Data Latih



Gambar 3.2 Diagram Alir Pengolahan Data Uji

3.6 Implementasi dan Analisa Sistem

Sehubungan dengan pengolahan data, maka pada tahap implementasi adalah tentang bagaimana pengolahan datanya diterapkan dalam sebuah sistem. Sistem yang dibuat diimplementasikan dengan menggunakan bahasa pemrograman Java Netbeans dan manajemen basis data MySQL. Penulis melakukan pengujian dan menganalisa sistem yang dihasilkan apakah sudah sesuai dengan yang diharapkan.

3.7 Penarikan Kesimpulan dan Penyusunan Laporan Tugas Akhir

Setelah dilakukan pengujian dan evaluasi terhadap sistem, maka selanjutnya adalah menarik kesimpulan dari hasil yang telah diperoleh. Setelah semua proses selesai dilakukan maka tahap terakhir adalah penyusunan laporan Tugas Akhir.



Gambar 3.3 Diagram Alir Metodologi Penelitian

BAB IV

ANALISIS DAN PERANCANGAN SISTEM

Pada bab ini akan dibahas mengenai analisis dan perancangan sistem. Pembahasan pada bab ini meliputi analisis kebutuhan perangkat lunak dan perancangan sistem, dimana pada perancangan sistem meliputi perancangan data, perancangan proses klasifikasi dengan *Naïve Bayes Classifier*, dan perancangan antar muka sistem.

4.1 Analisis Kebutuhan Perangkat Lunak

Merujuk pada tujuan umum penelitian Tugas Akhir ini adalah mengimplementasikan Metode *Naïve Bayes Classifier* untuk pemilihan jenis asuransi berdasarkan demografi calon pemegang polis.

Sesuai dengan tujuan Tugas Akhir diatas, maka dibuat kriteria-kriteria yang harus dimiliki perangkat lunak yang akan dibangun. Kriteria tersebut antara lain:

1. Perangkat lunak mampu membaca sumber data yang tersimpan dalam basis data.
2. Perangkat lunak mampu melakukan proses klasifikasi data demografi nasabah asuransi yang mengambil produk MLP, SPP, SK maupun MHP.
3. Perangkat lunak mampu memprediksi kelas produk nasabah asuransi pada data baru

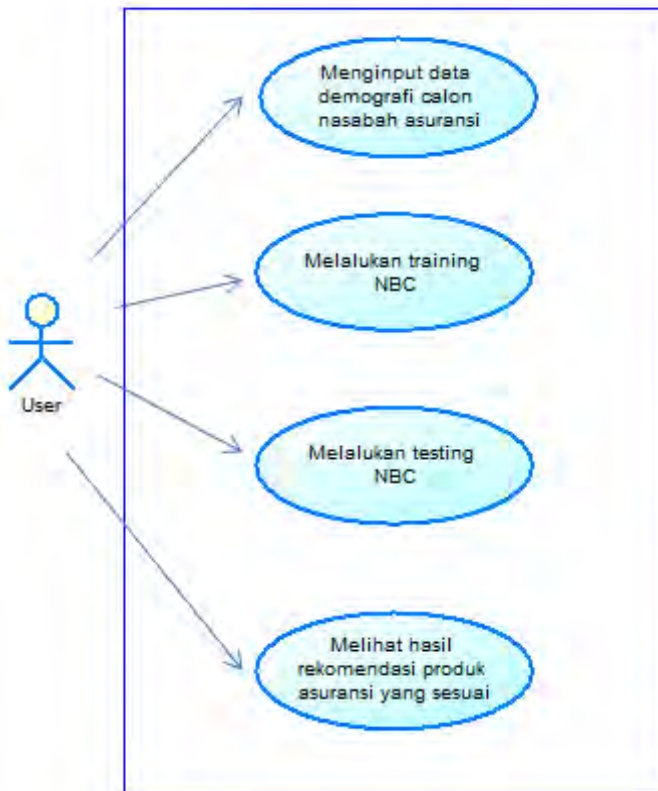
Untuk memenuhi kriteria tersebut, maka dilakukan langkah perancangan sistem.

4.2 Perancangan Sistem

Sistem yang dibuat nantinya akan melakukan perhitungan dan memberikan informasi dalam pemilihan jenis asuransi berdasarkan informasi demografi calon pemegang polis. Sistem ini menggunakan Metode *Naïve Bayes Classifier*.

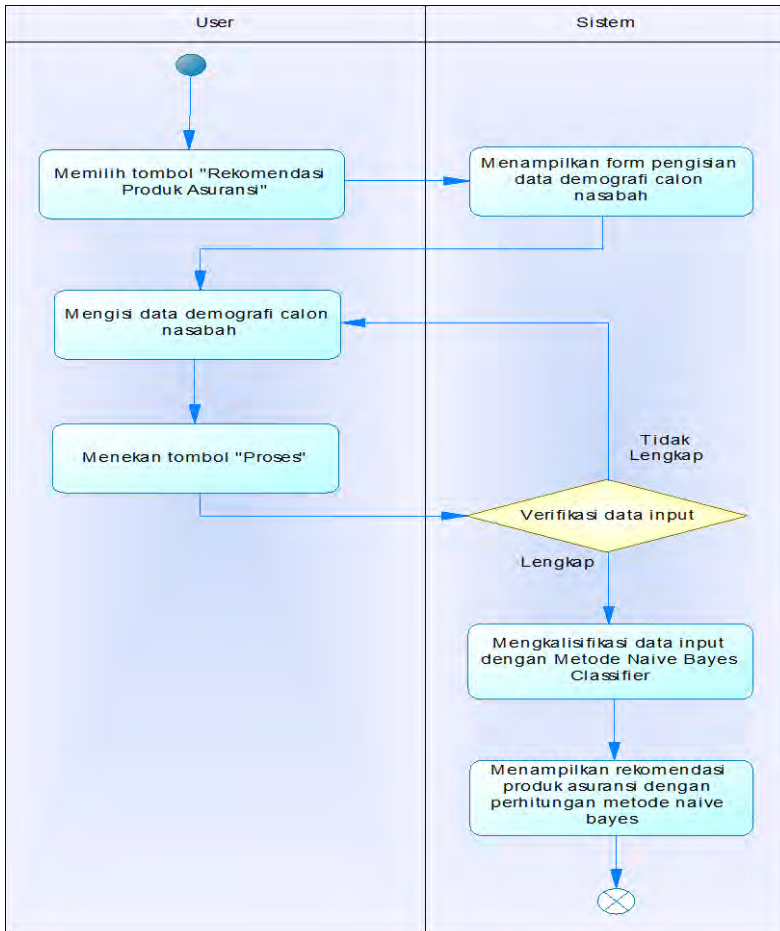
4.2.1 Pemodelan Sistem

Bab ini akan menjabarkan tentang *Use Case Diagram* dan *Diagram activity* perangkat lunak rekomendasi produk asuransi ini. *Use Case Diagram* dapat dilihat pada Gambar 4.1.



Gambar 4.1 *Use Case Diagram* Sistem Rekomendasi Asuransi

Setelah didefinisikan *Use Case* dari sistem, selanjutnya akan dibuat aktifitas diagram yang terjadi ketika *user* menggunakan sistem rekomendasi asuransi.



Gambar 4.2 Activity Diagram Sistem Rekomendasi Asuransi

4.2.2 Perancangan Data

Data yang digunakan dalam sistem ini nantinya akan dibagi ke dalam 3 kategori, yakni data masukan (*input*), data proses, dan data keluaran (*output*).

1. Data Masukan (*Input*)

Data *input* yang digunakan adalah data yang disimpan dalam tabel dari suatu basis data, dalam hal ini peneliti menggunakan Mysql. Data *input* merupakan data historis nasabah asuransi. Data tersebut memiliki 11 atribut yang terdiri dari produk asuransi, jenis kelamin, usia, pekerjaan, bidang pekerjaan, pendapatan, status pernikahan, jumlah anak, status merokok, dan riwayat kesehatan *pre-existing*. Data ini memiliki 518 *record* data nasabah asuransi. Dimana 117 data nasabah yang memilih produk asuransi MLP, 142 data nasabah yang memilih produk asuransi SPP, 108 data nasabah yang memilih produk asuransi jenis SK, dan 97 data nasabah yang memilih produksi asuransi MHP.

2. Data Proses

Data Proses adalah data yang digunakan oleh sistem selama proses berlangsung. Proses yang dimaksud adalah *preprocessing*, *training* dan *testing*. Data proses yang digunakan dalam penelitian ini antara lain: 8 data kategorik, 2 data numerik, dan 1 data kelas (target). Dimana atribut yang termasuk data kategorik antara lain;

- a. Jenis kelamin tertanggung asuransi yang dikelompokkan menjadi Perempuan Dan Laki-Laki,
- b. Pekerjaan tertanggung asuransi yang dikelompokkan menjadi Belum Bekerja, Mahasiswa/Pelajar, Karyawan, Ibu Rumah Tangga, TNI/Polri, PNS, Dan Pemilik Usaha.
- c. Bidang pekerjaan tertanggung asuransi yang dikelompokkan menjadi Jasa/Keuangan, Pemerintahan, Transportasi, Perdagangan, Konstruksi, Manufaktur, dan Sumber Daya Alam (SDA).
- d. Status Pernikahan tertanggung asuransi yang dikelompokkan menjadi Menikah, Belum Menikah, Dan Duda/Janda.
- e. Jumlah anak tertanggung asuransi adalah atribut yang masukannya bernilai numerik namun didiskritisasi menjadi 3 kelompok yaitu 'Tidak Ada' jika calon nasabah belum

memiliki anak, ‘Normal’ apabila calon nasabah memiliki 1 sampai 2 anak dan ‘Banyak’ jika calon nasbaah memiliki anak lebih dari 2.

- f. Status merokok tertanggung asuransi dikelompokkan menjadi 2 kelompok yaitu perokok dan bukan perokok.
- g. Kondisi *pre-existing* tertanggung asuransi yaitu ada tidaknya riwayat penyakit *pre-existing* yang sudah didefinisikan sebelumnya, yaitu jika calon nasabah memiliki salah satu atau lebih penyakit *pre-existing*, antara lain kanker, hepatitis B, hepatitis C, Diabetes Militus, stroke, ginjal, pencernaan akut, dan kolera maka calon nasabah dikelompokkan ke dalam kelas ‘Ada’, sedangkan jika tidak memenuhi kondisi *pre-existing* maka calon nasabah digolongkan ke dalam kelas ‘Tidak Ada’ kondisi *pre-existing*.

Atribut yang termasuk data numerik/kontinyu dalam penelitian ini adalah sebagai berikut:

- a. Usia tertanggung asuransi merupakan usia calon nasabah ketika pertama kali masuk asuransi
- b. Gaji pembayar premi merupakan pendapatan pembayar premi asuransi, dimana pendapatan pembayar premi harus lebih dari Rp. 500.000,-

Data kelas (target) dibagi menjadi 4 kelas sebagai berikut:

- a. MLP (*Maestrolink Plus*) merupakan kelas asuransi yang berfokus pada investasi dan dana pensiun nasabah asuransi.
- b. SPP (*Super Protection Plan*) merupakan kelas asuransi yang berfokus pada proteksi diri nasabah asuransi.
- c. MHP (*Maestro Hospital Plan*) merupakan kelas asuransi yang berorientasi pada kesehatan nasabah asuransi.
- d. SK (*Smart Kidz*) merupakan kelas asuransi yang memfokuskan pada biaya pendidikan nasabah asuransi. Syarat untuk mengambil asuransi ini adalah usia tertanggung harus kurang dari 21 tahun dan calon nasabah tersebut masih berstatus belum pekerja atau pelajar/mahasiswa.

3. Data Output (*Output*)

Terdapat 2 bagian data *output*, yang pertama data yang dihasilkan oleh proses *training* dan *testing* yaitu berupa nilai peluang bersyarat, nilai peluang *prior* dan tingkat akurasi klasifikasi. Sedangkan yang kedua adalah data hasil prediksi yaitu kelas produk asuransi. Kelas produk asuransi yang digunakan adalah kelas dengan peluang terbesar. Yaitu kelas MLP didapat jika peluang produk MLP lebih besar dibanding produk SPP, SK, dan MHP. Sedangkan produk SPP diperoleh jika peluang posterior produk SPP lebih besar daripada peluang produk MLP, SK, dan MHP. Dan produk MHP didapat jika peluang produk SK lebih besar dibanding peluang MLP, SPP, maupun MHP. Dan produk MHP diperoleh apabila peluang posterior dari produk MHP lebih besar dibanding peluang produk MLP, SPP, maupun SK.

4.2.3 Perancangan Proses Klasifikasi dengan Naïve Bayes Classifier

Proses Klasifikasi dengan *Naïve Bayes Classifier* untuk memprediksi kelas produk nasabah asuransi terbagi menjadi 3 tahap.

4.2.3.1 Pemrosesan Awal Data

Preprocessing atau pemrosesan awal data merupakan tahap pertama yang perlu dilakukan sebelum melakukan *mining* terhadap data. Dimana data yang sudah diperoleh akan di normalisasi dengan kata lain data akan dihilangkan *noise*-nya. Dalam Tugas Akhir ini *preprocessing* yang dibutuhkan adalah dimulai dengan menyelidiki apakah atribut numerik yang akan digunakan dalam proses klasifikasi mengikuti populasi yang berdistribusi normal. Pengujian sampel data menggunakan software Minitab 17, jika data yang akan dipakai berdistribusi normal, artinya data tersebut dapat dipakai dalam proses selanjutnya, namun jika tidak berdistribusi normal, maka dicari distribusi yang paling cocok dengan data tersebut serta mengganti probabilitas kondisionalnya

menjadi ungsi *pdf* dari distribusi data tersebut. Langkah selanjutnya adalah penghapusan duplikasi data, penanganan nilai yang hilang dan tranformasi data pada tipe data numerik yang di diskretisasi.

Tabel 4.1 Contoh *Dataset* Data Historis Nasabah Asuransi

No Polis	Gender	Usia	Pekerjaan	Gaji	Status Pernikahan	Status Merokok	Plan Asuransi
509-2639359	L	61	Karyawan	20	Menikah	Y	SPP
509-1165802	P	55	Pemilik Usaha	6	Menikah	T	MHP
509-1165877	P	8	Mahasiswa/Pelajar	11	Belum Menikah	T	MHP
509-1252030	P	53	Pemilik Usaha	10	Menikah	T	MHP
509-2639334	L	43	Karyawan	6	Menikah	Y	SPP
509-2638740	L	34	Karyawan	5	Belum Menikah	T	MLP
509-2180222	L	34	Karyawan	5	Menikah	Y	MLP
509-2559453	L		Karyawan	8.5	Menikah	T	SPP
509-2179364	L	24	Karyawan	5	Belum Menikah	T	MLP
308-9834349	P	33		10	Menikah	T	MLP
509-2623197	L	2	Belum Bekerja	10	Belum Menikah	T	SK
509-3004975	L	8	Mahasiswa/Pelajar	6	Belum Menikah	T	MHP
509-2880474	L	7	Mahasiswa/Pelajar	5	Belum Menikah	T	SPP
309-2559875	L	30	Pemilik Usaha	6	Menikah	T	MLP
509-1165935	P	28	Pemilik Usaha	7	Menikah	T	MHP
509-1038439	P	1	Belum Bekerja	5	Belum Menikah	T	SK
508-2055970	L	45	Karyawan	8	Menikah	Y	SPP
509-1038660	P	22	Karyawan	5	Belum Menikah	T	SPP
509-2639052	L	57	Pemilik Usaha	15	Menikah	Y	MLP
509-2956746	P	1	Belum Bekerja	10	Belum Menikah	T	SK
509-1038439	P	1	Belum Bekerja	5	Belum Menikah	T	SK

Contoh 4.1 Berdasarkan *dataset* yang terdapat pada Tabel 4.1, maka dapat dilakukan pemrosesan awal data sebelum dilakukan proses *Data Mining*.

Atribut ‘no polis’ merupakan *primary key* dalam *dataset* diatas, dapat diartikan bahwa satu No. polis kemunculannya tidak bisa lebih dari satu dalam suatu *dataset*. Pada Tabel 4.1, atribut no polis dengan No Polis=’509-1038439’ memiliki 2 *record*, artinya terjadi duplikasi data dalam *dataset* tersebut. Maka *record-record* yang sama tersebut dihapus dan disisakan 1 *record* dalam *dataset*. Dalam pemilihan data yang akan dihapus dapat dilihat dari kelengkapan nilai pada setiap atribut untuk masing-masing *record* yang memiliki nomor polis yang sama, jika ada nilai yang hilang

pada salah satu *record* maka *record* tersebutlah yang akan dihapus, tetapi jika kelengkapan nilai atribut sudah dipenuhi setiap *record*, maka data pertama sebagai data yang dianggap *valid* dan yang lain akan dihapus. Sehingga dari Tabel 4.1 dataset yang semula berjumlah 21 *record* kini menjadi 20 *record* setelah penghapusan *record* dengan “no polis” yang sama.

Kemudian, dalam Tabel 4.1 juga terlihat ada beberapa data yang memiliki nilai yang hilang. Oleh karena itu, nilai yang hilang tersebut harus ditangani agar pada saat proses *data mining* tidak ada *noise* dalam data. Ada beberapa cara konvensional yang biasa digunakan untuk menangani nilai yang hilang, yaitu dengan menghapus *record* yang memiliki *missing value* atau mengganti nilai yang hilang dengan nilai *default* yang sudah ditentukan. [8] Namun ada cara lain yang dapat digunakan, yaitu mengganti nilai yang hilang dengan modus untuk atribut kategorik dan mean untuk atribut kontinyu (numerik) [8]. Sehingga, untuk atribut ‘Pekerjaan’ dapat dicari frekuensi kemunculan untuk setiap kategori dalam atribut ‘Pekerjaan’.

Tabel 4.2 Frekuensi Kemunculan Tiap Atribut ‘Pekerjaan’

Pekerjaan	Frekuensi
karyawan	8
Pemilik Usaha	5
Mahasiswa/Pelajar	3
Belum Bekerja	3

Karena ‘Karyawan’ dalam atribut ‘Pekerjaan’ mempunyai frekuensi kemunculan terbanyak, maka untuk setiap nilai yang hilang pada atribut ‘Pekerjaan’ akan diganti dengan nilai ‘Karyawan’.

Selain itu, ada lagi data pada atribut ‘age’ yang memiliki *missing value*, karena ‘age’ merupakan atribut numerik, maka untuk mengatasinya bisa digunakan nilai rata-rata dari atribut ‘age’. Sehingga:

$$\begin{aligned} \text{mean}(\text{age}) &= \frac{61 + 55 + 8 + 53 + 43 + 34 + 34 + 24 + 33 + 2 + 8 + 7 + 30 + 28 + 1 + 45 + 22 + 57}{19} \\ &= \frac{546}{19} = 28.74 \approx 29 \end{aligned}$$

Hasil rata-rata usia dari *dataset* tersebut adalah 29 maka *record* yang memiliki nilai yang hilang pada atribut ‘age’ akan diganti dengan nilai 29.

Proses awal data yang dilakukan selanjutnya adalah transformasi data. Yaitu mentransformasi data sehingga data tersebut sesuai dengan bentuk yang diperlukan dalam proses *data mining*. Dalam hal ini, yang dilakukan dalam transformasi data adalah mengurangi atribut data yang tidak diperlukan dalam proses klasifikasi, yaitu atribut no polis. Sehingga setelah dilakukan *preprocessing*, maka data akan menjadi seperti Tabel 4.3.

Tabel 4.3 Hasil Pemrosesan Awal Data

Gender	Usia	Pekerjaan	Gaji	Status Pernikahan	Status Merokok	Plan Asuransi
L	61	Karyawan	20	Menikah	Y	SPP
P	55	Pemilik Usaha	6	Menikah	T	MHP
P	8	Mahasiswa/Pelajar	11	Belum Menikah	T	MHP
P	53	Pemilik Usaha	10	Menikah	T	MHP
L	43	Karyawan	6	Menikah	Y	SPP
L	34	Karyawan	5	Belum Menikah	T	MLP
L	34	Karyawan	5	Menikah	Y	MLP
L	29	Karyawan	8.5	Menikah	T	SPP
L	24	Karyawan	5	Belum Menikah	T	MLP
P	33	Karyawan	10	Menikah	T	MLP
L	2	Belum Bekerja	10	Belum Menikah	T	SK
L	8	Mahasiswa/Pelajar	6	Belum Menikah	T	MHP
L	7	Mahasiswa/Pelajar	5	Belum Menikah	T	SPP
L	30	Pemilik Usaha	6	Menikah	T	MLP
P	28	Pemilik Usaha	7	Menikah	T	MHP
P	1	Belum Bekerja	5	Belum Menikah	T	SK
L	45	Karyawan	8	Menikah	Y	SPP
P	22	Karyawan	5	Belum Menikah	T	SPP
L	57	Pemilik Usaha	15	Menikah	Y	MLP
P	1	Belum Bekerja	10	Belum Menikah	T	SK
P	1	Belum Bekerja	5	Belum Menikah	T	SK

4.2.3.2 Tahap *Training*

Setelah tidak ada *noise* dalam dataset, tahap yang selanjutnya adalah melakukan proses *training*. Tahapan ini terdiri dari beberapa langkah:

1. Mencari Peluang *Prior*.

Untuk mencari nilai peluang prior $P(Y)$ tiap kelas, dapat dicari dengan menghitung pecahan tiap data pelatihan yang dimiliki tiap kelas.

Contoh 4.2 lihat tabel 4.3 karena ada 20 *record dataset* dengan 6 *record* tergolong kelas ‘MLP’ dan ‘SPP’, sehingga $P(Y = MLP) = P(Y = SPP) = 0.3$. Sedangkan untuk yang masuk asuransi ‘SK’ berjumlah 3 *record*, maka $P(Y = SK) = 0.15$. Dan untuk yang tergolong kelas ‘MHP’ berjumlah 5 *record*, sehingga peluang prior dari MHP adalah $P(Y = MHP) = 0.25$.

2. Mencari nilai peluang bersyarat

a. Mencari Nilai Peluang Bersyarat untuk Atribut Kategorik.

Pada atribut kategori x_i , peluang bersyarat $P(X_i = x_i | Y = y)$, $i = 1, 2, 3, \dots$ dicari menurut pecahan data *training* pada kelas y yang memuat nilai atribut X_i .

Contoh 4.3 berdasarkan tabel dataset yang terdapat pada tabel 4.3, maka dapat dicari peluang bersyarat sebagai berikut:

Tabel 4.4 Tabel Peluang Atribut ‘Gender’

	MLP	SPP	SK	MHP
L	0.833333	0.833333	0.333333	0.2
P	0.166667	0.166667	0.666667	0.8

Tabel 4.5 Tabel Peluang Atribut ‘Status Pernikahan’

	MLP	SPP	SK	MHP
Menikah	0.666667	0.666667	0	0.6
Belum Menikah	0.333333	0.333333	1	0.4

Dari Tabel 4.4 dapat dilihat bahwa pada peluang bersyarat $P(X_i = \text{Menikah} | Y = SK) = 0$, ini menyebabkan adanya kemungkinan perhitungan peluang posterior bernilai 0, sehingga NBC tidak dapat

mengklasifikasikan data. Oleh karena itu, dibutuhkan satu metode untuk mengatasi permasalahan tersebut. Salah satunya yaitu menggunakan Teknik *Laplacian Smoothing*.

Teknik *laplacian smoothing* digunakan untuk mengatasi nilai probabilitas kondisional pada *Naïve Bayes Classifier* yang dapat bernilai 0. Cara yang digunakan pada teknik ini adalah dengan menambahkan angka 1 pada atribut yang nilai probabilitasnya kondisionalnya sama dengan 0. Persamaan di bawah ini menunjukkan perhitungan nilai probabilitas kondisional dengan Teknik *Laplacian Smoothing*.

$$P(X_i|Y) = \frac{1 + n(X_i, Y)}{|W| + n(Y)}$$

dimana,

$n(X_i, Y)$ = Jumlah *term* X_i yang ditemukan di seluruh data latih dengan kategori Y .

$n(Y)$ = Jumlah *term* di seluruh data latih dengan kategori Y .

$|W|$ = Jumlah seluruh *term* X_i dari seluruh data latih

Maka dengan menggunakan teknik *laplacian* pada perhitungan *Likelihood* yang bernilai 0, maka tabel probabilitas untuk atribut ‘Status Pernikahan’.

Tabel 4.6 Tabel Peluang Atribut ‘Status Pernikahan’
Menggunakan Teknik *Laplacian Smoothing*

	MLP	SPP	SK	MHP
Menikah	0.666667	0.666667	0.038462	0.6
Belum Menikah	0.333333	0.333333	1	0.4

Tabel 4.7 Tabel Peluang Atribut ‘Pekerjaan’ Menggunakan Teknik *Laplacian Smoothing*

	MLP	SPP	SK	MHP
Karyawan	0.666667	0.833333	0.043478	0.04
Pemilik Usaha	0.333333	0.038462	0.043478	0.6
Mahasiswa/Pelajar	0.038462	0.166667	0.043478	0.4
Belum Bekerja	0.038462	0.038462	1	0.04

Tabel 4.8 Tabel Peluang Atribut ‘Status Merokok’ Menggunakan Teknik *Laplacian Smoothing*

	MLP	SPP	SK	MHP
Y	0.333333	0.5	0.038462	0.04
T	0.666667	0.333333	1	1

b. Mencari Nilai Peluang Bersyarat untuk Atribut Numerik

Cara untuk mencari peluang bersyarat $P(X_i = x_i | Y = y)$ untuk atribut yang bersifat numerik atau kontinyu, diasumsikan bentuk tertentu distribusi peluang untuk variable kontinyu dan mengestimasi parameter distribusi menggunakan data latih. Distribusi *Gaussian* sering digunakan untuk merepresentasikan peluang bersyarat untuk atribut numerik. Distribusi ini dikarakterisasi dengan 2 parameter yaitu *mean*(μ) dan *varian*(σ^2). Untuk kelas y_j , peluang bersyarat untuk atribut x_i adalah:

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sigma_{ij}\sqrt{2\pi}} \exp -\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2} \dots\dots\dots (4.1)$$

Parameter μ_{ij} dapat diestimasi berdasarkan *mean* dari $X_i(\bar{x})$ untuk seluruh data latih yang dimiliki kelas y_j . Dengan cara yang sama σ_{ij}^2 dapat diestimasi dari *varian*(s^2) data latih yang dimiliki kelas y_j .

Contoh 4.4 lihat atribut ‘age’ yang ditunjukkan pada Tabel 4.3. *mean* dan *varian* untuk atribut ‘age’ yang memiliki kelas MLP.

$$\bar{x} = \frac{34+34+24+33+30+57}{6} = \frac{212}{6} = 35.33$$

$$s^2 = \frac{(-1.33)^2+(-1.33)^2+(-11.33)^2+(-27.5)^2+(-10.5)^2+(-12.5)^2}{6-1} = \frac{635.33}{5} = 127.0667$$

mean dan *varian* untuk atribut 'age' yang memiliki kelas SPP.

$$\bar{x} = \frac{61+43+29+7+45+22}{6} = \frac{207}{6} = 34.5$$

$$s^2 = \frac{(26.5)^2+(8.5)^2+(-5.5)^2+(-2.33)^2+(-5.33)^2+(21.67)^2}{6-1} = \frac{1827.5}{5} = 365.5$$

mean dan *varian* untuk atribut 'age' yang memiliki kelas SK.

$$\bar{x} = \frac{2+1+1}{3} = \frac{4}{3} = 1.33$$

$$s^2 = \frac{(0.67)^2+(-0.33)^2+(-0.33)^2}{3-1} = \frac{0.67}{2} = 0.33$$

mean dan *varian* untuk atribut 'age' yang memiliki kelas MHP.

$$\bar{x} = \frac{55+8+8+28+53}{5} = \frac{152}{5} = 30.4$$

$$s^2 = \frac{(24.6)^2+(-22.4)^2+(-22.4)^2+(-2.4)^2+(22.6)^2}{5-1} = \frac{2125.2}{4} = 531.3$$

Diberikan *record* dengan usia=24 tahun, maka dapat dihitung peluang kelas bersyarat sebagai berikut:

$$P(\text{age} = 24|MLP) = \frac{1}{11.3\sqrt{2\pi}} \exp - \frac{(24-35.33)^2}{2(127.067)} = 0.021$$

$$= 0.12$$

$$P(\text{age} = 24|SPP) = \frac{1}{19.12\sqrt{2\pi}} \exp - \frac{(24-34.5)^2}{2(365.5)} = 0.018$$

$$P(\text{age} = 24|SK) = \frac{1}{0.58\sqrt{2\pi}} \exp - \frac{(24-1.33)^2}{2(0.33)} \approx 0$$

$$P(\text{age} = 24|MHP) = \frac{1}{23.05\sqrt{2\pi}} \exp - \frac{(24-30.4)^2}{2(531.3)} = 0.017$$

Contoh 4.4 lihat atribut ‘gaji’ yang ditunjukkan pada tabel 4.3. *mean* dan *varian* untuk atribut ‘gaji’ yang memiliki kelas MLP.

$$\bar{x} = \frac{5+5+5+10+6+15}{6} = \frac{41}{6} = 6.83$$

$$s^2 = \frac{(-2.67)^2 + (-2.67)^2 + (-2.67)^2 + (2.33)^2 + (-1.67)^2 + (-7.33)^2}{6-1} = \frac{84.139}{5} = 16.67$$

mean dan *varian* untuk atribut ‘gaji’ yang memiliki kelas SPP.

$$\bar{x} = \frac{20+6+8.5+5+8+5}{6} = \frac{52.5}{6} = 8.75$$

$$s^2 = \frac{(11.25)^2 + (-0.25)^2 + (-3.75)^2 + (-3.75)^2 + (-0.75)^2 + (-2.75)^2}{6-1} = \frac{162.87}{5} = 32.58$$

mean dan *varian* untuk atribut ‘gaji’ yang memiliki kelas SK.

$$\bar{x} = \frac{10+10+5}{3} = \frac{25}{3} = 8.33$$

$$s^2 = \frac{(1.67)^2 + (1.67)^2 + (-3.33)^2}{3-1} = \frac{16.67}{2} = 8.33$$

mean dan *varian* untuk atribut ‘gaji’ yang memiliki kelas MHP.

$$\bar{x} = \frac{6+11+10+6+7}{5} = \frac{40}{5} = 8$$

$$s^2 = \frac{(-2)^2 + (-1)^2 + (-2)^2 + (3)^2 + (2)^2}{5-1} = \frac{22}{4} = 5.5$$

Diberikan *record* dengan gaji=7 juta, maka dapat dihitung peluang kelas bersyarat sebagai berikut:

$$P(\text{gaji} = 7 | \text{MLP}) = \frac{1}{4.1\sqrt{2\pi}} \exp - \frac{(7-6.83)^2}{2(16.83)} = 0.097$$

$$P(\text{gaji} = 7 | \text{SPP}) = \frac{1}{5.71\sqrt{2\pi}} \exp - \frac{(7-8.75)^2}{2(32.575)} = 0.067$$

$$P(\text{gaji} = 7 | \text{SK}) = \frac{1}{2.89\sqrt{2\pi}} \exp - \frac{(7-8.33)^2}{2(8.33)} = 0.124$$

$$P(\text{gaji} = 7 | \text{MHP}) = \frac{1}{5.5\sqrt{2\pi}} \exp - \frac{(7-8)^2}{2(5.5)} = 0.155$$

4.2.3.3 Tahap Klasifikasi

Tahap ketiga yaitu tahap klasifikasi. Setelah model prediksi telah dibangun pada data latih, saatnya mengklasifikasi data yang belum diketahui label kelasnya.

Langkah-langkah pada tahap ini adalah:

1. Penentuan peluang bersyarat $P(X_i = x_i | Y = y)$ sesuai dengan hasil *training* dimana x_i sesuai dengan *input user*.
2. Perhitungan $\prod_{i=1}^d P(X_i = x_i | Y = y)$
3. Perhitungan peluang posterior $P(Y|X) = \frac{P(Y) \prod_{i=1}^d P(X_i|Y)}{P(X)}$
4. Untuk mengklasifikasi *record*, ada beberapa kondisi yang harus diperhatikan.
 - a. jika $P(MLP|X) > P(SPP|X), P(MLP|X) > P(SK|X), P(MLP|X) > P(MHP|X)$, maka *record* diklasifikasikan masuk kelas MLP.
 - b. Jika $P(SPP|X) > P(MLP|X), P(SPP|X) > P(SK|X), P(SPP|X) > P(MHP|X)$, maka *record* akan diklasifikasikan masuk kelas SPP.
 - c. Jika $P(MHP|X) > P(SPP|X), P(MHP|X) > P(SK|X), P(MHP|X) > P(MLP|X)$, maka *record* akan diklasifikasikan masuk kelas MHP.
 - d. Jika $P(SK|X) > P(SPP|X), P(SK|X) > P(MHP|X), P(SK|X) > P(MLP|X)$, maka ia tergolong dalam kelas SK. Namun, jika peluang posterior maksimum berada pada kelas SK tetapi usia nasabah lebih dari 21 tahun atau sudah bekerja, maka nasabah tersebut tidak memenuhi syarat untuk mengambil produk asuransi SK, sehingga ia akan direkomendasikan pada kelas dengan peluang posterior maksimum kedua.

Contoh 4.6 untuk memprediksi label kelas pada data uji (*test record*) berikut:

* Gender=L * Pekerjaan=Karyawan * SM=Perokok
 * Usia=24 th * Status=Menikah * Gaji=7 jt/bln

Tahap pertama, penentuan peluang kelas bersyarat $P(X_i = x_i | MLP), P(X_i = x_i | SPP), P(X_i = x_i | SK)$,

$P(X_i = x_i|MHP)$ yang dihasilkan dari proses *training*. Lihat Contoh 4.3 dan 4.4, maka peluang bersyarat yang digunakan adalah:

Peluang bersyarat atribut ‘gender=L’

	MLP	SPP	SK	MHP
L	0.8333	0.8333	0.3333	0.2

Peluang bersyarat atribut ‘Pekerjaan=Karyawan’

	MLP	SPP	SK	MHP
Karyawan	0.6667	0.8333	0.0434	0.04

Peluang bersyarat atribut ‘Status Pernikahan=Menikah’

	MLP	SPP	SK	MHP
Menikah	0.6667	0.6667	0.038462	0.6

Peluang bersyarat atribut ‘Status Merokok= Perokok’

	MLP	SPP	SK	MHP
Perokok	0.6667	0.6667	0.038462	0.6

Peluang bersyarat atribut ‘Gaji= 7’

	MLP	SPP	SK	MHP
Gaji=7	0.097	0.067	0.124	0.155

Peluang bersyarat atribut ‘Usia=24’

	MLP	SPP	SK	MHP
Usia=24	0.021	0.018	0	0.017

Tahap kedua, mencari $\prod_{i=1}^d P(X_i = x_i|Y = y)$ dari hasil yang diperoleh dari tahap pertama. $X' = (gender = L, age = 24, gaji = 7, S.M = Perokok, S.Pernikahan = Menikah)$.

$$\begin{aligned}
 P(X|MLP) &= (0.83) * (0.67) * (0.67) * (0.67) * (0.097) * \\
 &\quad (0.021) \\
 &= 0.00051
 \end{aligned}$$

$$\begin{aligned}
 P(X|SPP) &= (0.83) * (0.83) * (0.67) * (0.67) * (0.067) * \\
 &\quad (0.018) \\
 &= 0.000373
 \end{aligned}$$

$$\begin{aligned}
 P(X|SK) &= (0.33) * (0.43) * (0.039) * (0.039) * (0.124) * \\
 &\quad (0) \\
 &= 0
 \end{aligned}$$

$$\begin{aligned}
 P(X|MHP) &= (0.2) * (0.04) * (0.6) * (0.6) * (0.155) * \\
 &\quad (0.017) \\
 &= 7.59 \times 10^{-6}
 \end{aligned}$$

Tahap ketiga, menghitung peluang posterior $P(MLP|X)$, $P(SPP|X)$, $P(SK|X)$, $P(MHP|X)$ dengan $P(MLP) = 0.3$, $P(SPP) = 0.3$, $P(SK) = 0.15$, $P(MHP) = 0.25$, sehingga peluang posterior masing-masing kelas dapat dihitung dengan menggunakan rumus (2.3), maka didapat:

$$P(MLP|X) = (0.3) * (0.00051) = 0.000153$$

$$P(SPP|X) = (0.3) * (0.000373) = 0.000112$$

$$P(SK|X) = (0.15) * 0 = 0$$

$$P(MHP|X) = (0.25) * (7.59 \times 10^{-6}) = 1.9 \times 10^{-6}$$

Tahap ke empat, karena $P(MLP|X)$ memiliki nilai posterior paling besar daripada $P(SPP|X)$, $P(SK|X)$, dan $P(MHP|X)$, maka *record* diklasifikasikan dalam kelas MLP.

4.3 Perancangan Antar Muka Sistem

Perancangan antarmuka meliputi perancangan struktur menu dan perancangan tampilan pada tampilan *user*.

4.3.1 Perancangan Struktur Menu User

Pembuatan perangkat lunak klasifikasi jenis asuransi dengan Metode *Naïve Bayes Classifier* menggunakan struktur menu pada Gambar 4.1.

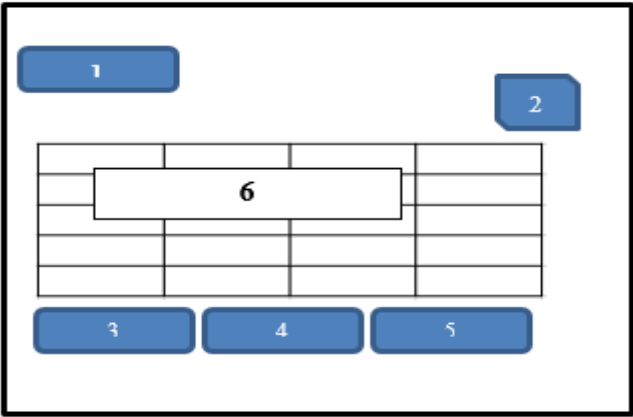


Gambar 4.3 Struktur Menu *User*

4.3.2 Perancangan Antar Muka Proses *Data Mining*

Bab ini menjelaskan tentang perancangan antar muka yang akan diimplementasikan ke sistem. Antar muka yang akan di rancang meliputi 2 *form* utama, yaitu tampilan *form* data mining dan *form* rekomendasi untuk data calon nasabah.

a. *Form Data Mining*

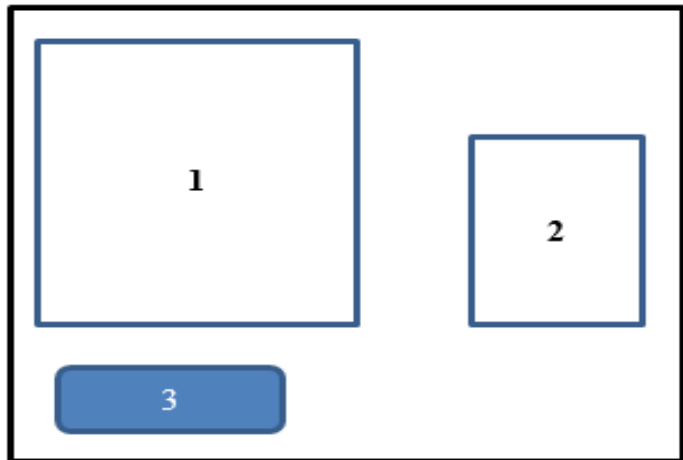


Gambar 4.4 Perancangan antar muka *Form Data Mining*

Keterangan Gambar 4.2:

1. Tombol 1 untuk mengarahkan *user* ke *form* operasi *database* nasabah asuransi
2. Kolom 2 pemilihan proporsi data latih. Dimana dalam item ini, *user* akan diberikan beberapa pilihan proporsi data latih yang akan digunakan dalam proses pelatihan
3. Tombol 3 untuk melakukan pemrosesan awal data (*preprocessing*). Tombol ini akan segera aktif setelah *user* menampilkan data historis nasabah asuransi.
4. Tombol 4 untuk melakukan pelatihan pada data latih. Tombol ini akan segera diaktifkan ketika *preprocessing* selesai dilakukan
5. Tombol 5 untuk melakukan pengujian pada data uji. Ketika menekan tombol ini *user* diarahkan menuju *form* untuk melakukan pengujian terhadap data uji dan melihat kinerja sistem yang telah dihasilkan
6. Tabel 6 tempat menampilkan *dataset* nasabah asuransi.

b. *Form* Rekomendasi Data Baru



Gambar 4.5 Perancangan Antar Muka Form Rekomendasi Calon Nasabah

Keterangan Gambar 4.3:

1. Kolom 1 digunakan untuk mengisi data demografi calon nasabah. *User* bisa melengkapi profil demografi calon nasabah asuransi yang nanti akan diberikan rekomendasi tentang pemilihan jenis asuransi.
2. Kolom 2 menampilkan hasil rekomendasi produk asuransi bagi calon nasabah. Kolom ini akan menampilkan urutan jenis-jenis asuransi bagi calon nasabah berdasarkan ciri demografi yang sudah dimasukkan menurut ranking probabilitas posterior masing-masing produk asuransi.
3. Tombol 3 digunakan untuk melakukan klasifikasi dengan Metode *Naïve Bayes Classifier*. Tombol ini akan menginstruksikan sistem untuk melakukan perhitungan posterior berdasarkan data demografi calon nasabah yang sudah dimasukkan.

BAB V

IMPLEMENTASI DAN ANALISIS SISTEM

Dalam bab ini dijelaskan tentang implementasi dan analisis dari hasil uji coba sistem dengan parameter yang berbeda.

5.1 Implementasi Sistem

Pada sub bab ini akan diuraikan tentang implementasi perangkat lunak yang meliputi uraian tentang modul-modul utama sistem dan implementasi antar muka. Dalam pembuatan perangkat lunak ini, Bahasa yang digunakan bahasa java netbeans, sedangkan sistem basis data yang digunakan untuk implementasi sistem ini adalah Mysql.

Perangkat lunak diimplementasikan dengan menggunakan komputer dengan sistem operasi Microsoft Windows 8 dengan spesifikasi perangkat keras prosesor Intel® Coleron® CPU 1005M @1.90GHz (2CPUs), ~1.9GHz dengan memori utama sebesar 2 GB. Spesifikasi tersebut digunakan sebagai standar aplikasi dan pengujian perangkat lunak.

5.1.1 Implementasi Kode Program (*Source Code*)

Pada sub bab 4.4 telah dijelaskan tentang perancangan sistem perangkat lunak yang terdiri dari beberapa proses. Berikut akan diuraikan modul-modul utama yang digunakan dalam implementasi sistem.

a. Modul Koneksi *Database*

‘Fungsi koneksi *database*’

```
public void koneksi(){
    try {

        Class.forName("com.mysql.jdbc.Driver");
        conn=DriverManager.getConnection("jdbc:mysql://
        localhost:3306/asuransidb","root","");
        cn=conn.createStatement();
    }catch(Exception e){
```

```
JOptionPane.showMessageDialog(null, "koneksi
gagal");
        System.out.println(e.getMessage());
    }}
```

b. Modul Proses *Training*

Pada proses *training*, terdapat serangkaian langkah yang tentunya memiliki modul tersendiri. Berdasarkan metodologi penelitian, proses *training* terdiri dari *training*, *testing* dan perhitungan presisi serta akurasi. Untuk mencari nilai peluang prior tiap kelas ($P(Y)$), dapat dicari dengan menghitung pecahan tiap *training record* yang dimiliki tiap kelas.

‘Fungsi untuk menghitung peluang prior’

```
sql = "SELECT * FROM nasabah";
stat = conn.createStatement();
res = stat.executeQuery(sql);
res.last();
jum = res.getRow();
res.beforeFirst();
sql = "SELECT nama_produk FROM produktb";
stat = conn.createStatement();
res = stat.executeQuery(sql);
while(res.next()){
    String sql1="select no_polis from
nasabah where
plan_asuransi='"+res.getString("nama_produk")+"'
";
    ResultSet r=
conn.createStatement().executeQuery(sql1);
    r.last();
    tot=r.getRow();
    r.beforeFirst();
    op=(double) tot/jum;
    String sql2="UPDATE produktb SET
peluang_produk= "+op+" WHERE nama_produk='"+
res.getString("nama_produk") + "'";
```

Mencari peluang bersyarat untuk atribut kategorik, pada atribut kategori x_i , peluang bersyarat $P(X_i = x_i | Y = y)$ dicari menurut pecahan data latih pada kelas Y yang memuat nilai atribut x_i . Agar nilai *likelihood* tidak menghasilkan nilai 0 yang memungkinkan Metode *Naïve Bayes Classifier* tidak dapat mengklasifikasikan suatu *record*, maka pada tahap perhitungan nilai *likelihood* digunakan Teknik *Laplacian Smoothing* pada atribut yang memiliki probabilitas kondisional bernilai 0.

'Fungsi menghitung peluang bersyarat tiap atribut kategorik'

```
public void hitungPeluang(ResultSet res, String
atr){
    String sql3="select jenis from
kategoritb where atribut='"+atr+"'";
    ResultSet
s=conn.createStatement().executeQuery(sql3);
    while(s.next()){
        String sql4="select * from
nasabah where "+atr+"='"+s.getString("jenis")+"'
and
plan_asuransi='"+res.getString("nama_produk")+"'
";
        ResultSet k=
conn.createStatement().executeQuery(sql4);
        k.last();
        int ad=k.getRow();
        k.beforeFirst();
        String d="select no_polis from nasabah where
"+atr+"='"+s.getString("jenis")+"'";
        ResultSet
rl=conn.createStatement().executeQuery(d);
        rl.last();
        int at=rl.getRow();
        rl.beforeFirst();
        opl= (double) ad/tot;
        if(opl==0){
            opl=(double) (1+ad)/(at+tot);
        }
    }
}
```

Cara untuk mencari peluang bersyarat $P(X_i = x_i | Y = y)$ untuk atribut numerik (kontinyu). Distribusi *Gaussian* sering digunakan untuk merepresentasikan peluang bersyarat untuk atribut kontinyu. Distribusi *Gaussian* dikarakteristikan dengan 2 parameter yaitu *mean* (μ) dan *Varian* (σ^2).

‘Fungsi menghitung Mean tiap kelas’

```
String sql7="select "+str+" from nasabah where
plan_asuransi='"+res.getString("nama_produk")+"'
";
        ResultSet k=
conn.createStatement().executeQuery(sql7);
        double n=0;
        while(k.next()){
            n=n+k.getDouble(str);
        }
```

‘Fungsi menghitung Varian tiap kelas’

```
String sql7="select "+str+" from nasabah where
plan_asuransi='"+res.getString("nama_produk")+"'
";
        ResultSet
k=conn.createStatement().executeQuery(sql7);
double mean=(double) n/tot;
        System.out.println("mean= "+mean);

k=conn.createStatement().executeQuery(sql7);
        n=0;
        while(k.next()){
            n=n+Math.pow(k.getDouble(str)- mean, 2);
        }
        double SD=(double) (n/tot);
```

‘Fungsi menghitung Distribusi Gaussian’

```
String sql="select * from probabilitastb where
atribut='"+at+"' and
jenis_produk='"+rt.getString("nama_produk")+"'";
        ResultSet l=
conn.createStatement().executeQuery(sql);
        while(l.next()){
            m=l.getDouble("mean");
```

```

sd=l.getDouble("standar_dev");
double var=Math.sqrt(sd);
double eq1= 1/var*(Math.sqrt(2*Math.PI));
double eq3=Math.pow((k-m),2)/(2*sd);
double eq2=1/Math.exp(eq3);
p1=eq1*eq2;

```

c. Modul Klasifikasi

Klasifikasi merupakan proses penentuan kelas dari instance yang belum diketahui kelas sebelumnya. Dalam Naïve Bayes Classifier akan dihasilkan kelas produk rekomendasi asuransi MLP, SPP, SK, atau MHP. Kelas rekomendasi asuransi yang digunakan adalah kelas yang memiliki peluang posterior maksimum.

‘Fungsi menentukan kelas klasifikasi’

```

String sql1="select * from produktb";
ResultSet
rt=conn.createStatement().executeQuery(sql1);
while(rt.next()){
    double
    Prob=rt.getDouble("peluang_produk");

    System.out.println(rt.getString("nama_produk")
    );
    procal=Prob*nilai("children",
    res.getString("children"),
    rt)*gaussian("salary",
    res.getDouble("salary"), rt)*gaussian("age",
    res.getDouble("age"), rt)*nilai("gender",
    res.getString("gender"), rt)*nilai("job",
    res.getString("job"), rt)*nilai("field",
    res.getString("field"),
    rt)*nilai("marital_status",
    res.getString("marital_status"),
    rt)*nilai("smoker_status",
    res.getString("smoker_status"),
    rt)*nilai("medical_hist",
    res.getString("medical_hist"), rt);
}

```



```

        if(res.getInt("age")>=22 || !"Belum
Bekerja".equals(res.getString("job"))){

        if(!"Mahasiswa/Pelajar".equals(res.getString("
job"))){

                                String sql2="UPDATE
produktb SET peluang_uji="+0+" WHERE
nama_produk='SK'";

        cn.executeUpdate(sql2);
        }

        String sql3="select
max(peluang_uji) from produktb";
        ResultSet
t=conn.prepareStatement(sql3).executeQuery();

```

5.1.2 Implementasi Data

Pada sub bab 4.2.1 telah dirancang data yang akan dipakai pada perangkat lunak. Berikut adalah implementasi data tersebut. Struktur data dari tabel nasabah ditampilkan dalam tabel 5.1. pada Tabel 5.1, ditunjukkan pada tabel nasabah terhadap 10 field, 9 field diantaranya 'gender', 'age', 'job', 'field', 'salary', 'marital_status', 'children', 'smoker_status', 'medical_hist' adalah *predictor variable* dan plan_asuransi adalah *predicted variable*.

Tabel 5.1 Tabel Nasabah Terhadap 10 Atribut.

Nama atribut	Tipe	Keterangan
gender	varchar	Jenis kelamin nasabah
age	int	Usia nasabah
job	varchar	Pekerjaan nasabah
field	varchar	Bidang perkerjaan nasabah

salary	double	Gaji pembayar premi asuransi/bulan (juta)
Marital_status	varchar	Status perkawinan nasabah
children	varchar	Jumlah anak nasabah yang sudah dideskritisasi
Smoker_status	varchar	Status merokok nasabah
Medical_hist	varchar	Riwayat penyakit <i>pre-existing</i> nasabah
Plan_asuransi	varchar	Produk asuransi nasabah

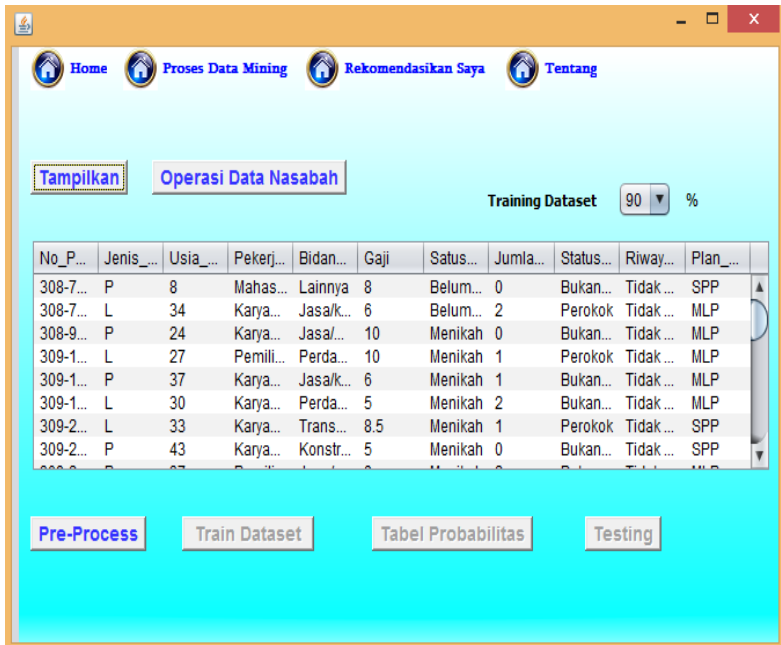
5.1.3 Implementasi Antar Muka

Untuk implementasi antar muka terhadap menu utama, yaitu Home, Data Mining, Rekomendasikan Saya, dan Tentang.

a. Menu Data Mining

Tampilan utama yang penting dalam menu ini adalah *form training*, *form* menampilkan tabel probabilitas hasil *training*, dan halaman *testing*. Untuk memulai proses, *user* harus masuk pada menu '*Data Mining*'. Dalam menu ini, *user* bisa menampilkan data historis polis nasabah asuransi, melakukan *preprocessing*, mengatur perbandingan *instance* yang akan dipakai untuk proses *training*, melakukan proses *training*, melihat tabel probabilitas dan melakukan testing pada data uji.

Sebelum melakukan pemrosesan awal data atau *preprocessing*, *user* diharuskan menampilkan data historis nasabah asuransi.



Gambar 5.1 Tampilan Sistem Untuk Menampilkan Data Historis

Setelah data historis ditampilkan, tombol ‘*pre-process*’ segera diaktifkan sehingga *user* bisa melakukan *preprocessing* pada data historis nasabah asuransi. Setelah melakukan pemrosesan awal pada data historis nasabah, *user* harus memilih perbandingan data latih yang nantinya akan digunakan dalam proses *training* dengan memilih pilihan angka pada *ComboBox* di ujung kanan atas. Dalam contoh berikut akan digunakan 90% data latih.

Kemudian, setelah melakukan proses *training*, sistem akan menghitung semua nilai probabilitas kondisional dari masing-masing atribut. Setelah dilakukan proses *training*, *user* juga dapat melihat tabel probabilitas yang sudah dihasilkan dari proses *training* seperti Gambar 5.2.

Home Proses Data Mining Rekomendasi Saya Tentang Close

Tabel Peluang "Jenis Kelamin"

Katego...	Plan As...	Peluang
L	MLP	0.574
L	SPP	0.606
L	SK	0.527
L	MHP	0.424
P	MLP	0.426
P	SPP	0.394
P	SK	0.473
P	MHP	0.576

Tabel Peluang "Status Pernikahan"

Katego...	Plan A...	Peluang
Belum...	MLP	0.213
Belum...	MHP	0.391
Belum...	SPP	0.465
Belum...	SK	0.989
Duda...	MHP	0.163
Duda...	MLP	0.045
Duda...	SPP	0.008
Duda...	SK	0.009

Tabel Peluang "Pekerjaan"

Katego...	Plan A...	Peluang
Belum...	MLP	0.006
Belum...	MHP	0.163
Belum...	SK	0.559
Belum...	SPP	0.039
Ibu Ru...	SK	0.01
Ibu Ru...	SPP	0.016
Ibu Ru...	MLP	0.006
Ibu Ru...	MHP	0.011

Mean dan Varian "Usia"

Plan As...	Mean	Varian
MHP	27.826	340.252
MLP	35.987	118.439
SK	4.656	14.398
SPP	27.126	206.236

Tabel Probabilitas "Jumlah Anak"

Plan A...	Plan A...	Peluang
Banyak	MHP	0.12
Normal	MHP	0.391
Tidak ...	MHP	0.489
Tidak ...	MLP	0.29
Banyak	MLP	0.148
Normal	MLP	0.561
Normal	SK	0.004
Banyak	SK	0.007

Tabel Peluang "Bidang Pekerjaan"

Kategor...	Plan A...	Peluang
JasaK...	MLP	0.574
JasaK...	SPP	0.079
JasaK...	MHP	0.196
JasaK...	SK	0.005
Konstr...	SPP	0.22
Konstr...	SK	0.008
Konstr...	MLP	0.005
Konstr...	MHP	0.011

Tabel Peluang "Status Merokok"

Kategor...	Plan As...	Peluang
T	MLP	0.645
T	SPP	0.638
T	SK	1
T	MHP	0.946
Y	MLP	0.365
Y	SPP	0.362
Y	SK	0.005
Y	MHP	0.054

Tabel Peluang "Riwayat Penyakit"

Kategor...	Plan As...	Peluang
A	MLP	0.013
A	SPP	0.008
A	SK	0.008
A	MHP	0.315
T	MLP	0.987
T	SPP	0.992
T	SK	1
T	MHP	0.674

Mean dan Varian "Gaji Pemegang Rekening"

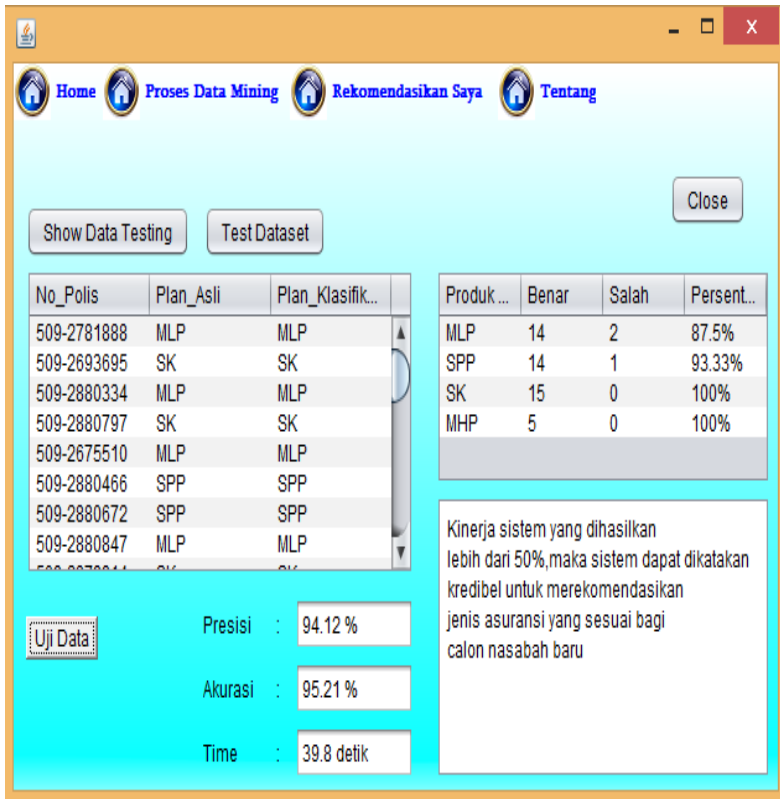
Plan As...	Mean	Varian
MHP	8.232	16.523
MLP	7.034	20.851
SK	6.959	6.23
SPP	10.537	76.071

Tabel Peluang "Produk Asuransi"

Produk Asura...	Peluang
MLP	0.332
SPP	0.272
SK	0.199
MHP	0.197

Gambar 5.2 Tabel Probabilitas Yang Dihasilkan dari Proses *Training*

Kemudian, tombol pada proses *testing* telah aktif, artinya *user* sudah dibolehkan untuk melihat hasil *testing* pada data uji. Dalam contoh kali ini, proporsi data uji yang dipakai adalah 10% dari data historis nasabah asuransi.



Gambar 5.3 Hasil Pengujian dari Data Uji

b. Menu Rekomendasi Baru

Setelah melakukan proses *training* dan proses *testing*, akan diperoleh nilai presisi dan akurasi dari sistem yang telah dihasilkan. Jika nilai kinerja sistem $\geq 50\%$ artinya probabilitas kondisional yang dihasilkan pada proses *training* dapat digunakan untuk melakukan rekomendasi. Untuk melakukan rekomendasi, klik menu 'Rekomendasi Saya' kemudian akan muncul *form* pengisian data calon nasabah untuk rekomendasi produk asuransi seperti Gambar 5.4.

Home Proses Data Mining Rekomendasi Saya Tentang

Profil Tertanggung

Jenis Kelamin : Laki-Laki

Tanggal Lahir : 2 Feb 1979

Pekerjaan : Karyawan

Bidang Pekerjaan : Konstruksi

Status Pernikahan : Menikah

Jumlah Anak : 3

Status Merokok : Bukan Perokok

Kondisi Pre-Existing : ☒ Tidak Ada

**Apabila anda memiliki salah satu penyakit pre-existing dibawah ini, maka pilih button 'Ada'*

Kanker Hepatitis B Hepatitis C Diabetes Milius Stroke

Gigit Pencemaran Akut Kolera

Profil Pembayar Premi

Jenis Kelamin : Laki-Laki

Pekerjaan : Karyawan

Bidang Pekerjaan : Konstruksi

Gaji : 8 juta

Rekomendasikan Saya

Reset

Waktu Rekomendasi :

**Apabila anda memenuhi kondisi pre-existing*

Anda diharuskan melakukan medical tes untuk penerimaan asuransi

Gambar 5.4 Form Rekomendasi Untuk Input Data Baru

Dalam pengujian Gambar 5.4, digunakan contoh data calon tertanggung adalah seorang laki-laki yang lahir pada tanggal 2 februari 1979, ia bekerja sebagai seorang karyawan di bidang konstruksi, dengan status pernikahan ‘menikah’ dan memiliki 3 orang anak serta bukan seorang perokok dan tidak ada riwayat penyakit *pre-existing*. Sementara penghasilan rata-rata perbulan dari pembayar premi asuransi sebesar 8 juta.

Setelah data lengkap dimasukkan, kemudian tekan *button* ‘Rekomendasikan Saya’, maka proses rekomendasi produk

asuransi dengan menggunakan Metode *Naïve Bayes Classifier* akan dijalankan dan hasilnya seperti Gambar 5.5:

Home Proses Data Mining Rekomendasikan Saya Tentang

Profil Tertanggung

Jenis Kelamin : Laki-Laki

Tanggal Lahir : 6 Feb 1979

Pekerjaan : Karyawan

Bidang Pekerjaan : Konstruksi

Status Pernikahan : Menikah

Jumlah Anak : 3

Status Merokok : Bukan Perokok

Kondisi Pre-Existing : ☒ Tidak Ada ☐ Ada

**Apabila anda memiliki salah satu penyakit pre-existing dibawah ini, maka pilih button 'Ada'*

Kanker Hepatitis B Hepatitis C Diabetes Militus Stroke
Ginjal Pencernaan Akut Kolera

Profil Pembayar Premi

Jenis Kelamin : Laki-laki

Pekerjaan : Karyawan

Bidang Pekerjaan : Konstruksi

Gaji : 8 jt

Rekomendasikan Saya

Reset

Waktu Rekomendasi : 0.34 detik

**Apabila anda memenuhi kondisi pre-existing
Anda diharuskan melakukan medical tes untuk penerimaan asuransi*

Sebaiknya anda memilih asuransi dengan urutan sebagai berikut

1. SPP
SPP (Super Protection Plan) Merupakan
2. MLP
MLP (MaestroLink Plus) Merupakan asur
3. MHP
MHP (Maestro Hospital Plan) merupakan

Gambar 5.5 Hasil Rekomendasi Data Baru

5.2 Uji Coba Sistem

Dalam bab ini dijelaskan hasil uji coba perangkat lunak. Data yang akan dipakai adalah data historis nasabah asuransi AXA Financial Indonesia selama kurun waktu 2 tahun yakni mulai januari 2014-januari 2016.

5.2.1 Lingkungan Uji Coba

Perangkat keras yang dipakai untuk uji coba sistem ini adalah komputer dengan sistem operasi *Microsoft Windows 8* dengan spesifikasi perangkat lunak prosesor Intel® Coleron® CPU 1005M @1.90GHz (2CPUs), ~1.9GHz dengan memori utama sebesar 2 GB.

5.2.2 Spesifikasi Dataset

Dataset yang digunakan memiliki 9 atribut input dan 1 atribut output. Pada atribut input terdapat 3 atribut numerik dan 6 atribut kategorikal. Sedangkan atribut *output* merupakan atribut kategori yang memiliki yang mempunyai kelas MLP yang berarti nasabah tersebut dianjurkan untuk mengambil produk tersebut, kelas SPP yang berarti nasabah tersebut direkomendasikan untuk mengikuti asuransi SPP, kelas SK menunjukkan bahwa nasabah disarankan untuk mengambil produk asuransi SK dan kelas MHP berarti nasabah dianjurkan untuk mengambil asuransi MHP sebagai prioritas utama. Dataset ini memiliki 518 *instance* nasabah historis asuransi.

5.3 Evaluasi Sistem

Uji coba dilakukan terhadap tabel nasabah data testing, berikut evaluasi yang telah dilakukan.

5.3.1 Jenis-jenis Pengamatan

Tujuan fungsionalitas aplikasi ini menguji kemampuan perangkat lunak dalam melakukan klasifikasi data testing.

Pengujian tingkat presisi dapat dirumuskan sebagai berikut:

$$\text{Presisi} = \frac{\text{jumlah instance benar}}{\text{total instance}} \times 100\%$$

Sedangkan yang kedua adalah pengamatan total akurasi yang dapat dirumuskan sebagai berikut:

$$\text{Akurasi} = \frac{\sum_{i=0}^d \text{prosentase benar kelas ke } - i}{\text{total kelas klasifikasi}}$$

5.3.2 Evaluasi Uji Coba

Untuk pengujian yang akan diamati adalah kemampuan melakukan rekomendasi dan hasil presisi serta akurasi sistem yang dihasilkan.

Uji coba untuk melakukan rekomendasi produk SPP sebagai pilihan utama seperti dibawah ini.

Profil Tertanggung

Jenis Kelamin : Laki-Laki

Tanggal Lahir : 2 Feb 1979

Pekerjaan : Karyawan

Bidang Pekerjaan : Konstruksi

Status Pernikahan : Menikah

Jumlah Anak : 3

Status Merokok : Bukan Perokok

Kondisi Pre-Existing : ☒ Tidak Ada ☐ Ada

**Apabila anda memiliki salah satu penyakit pre-existing dibawah ini, maka pilih button 'Ada'*

Kanker Hepatitis B Hepatitis C Diabetes Melitus Stroke
Ginjal Pencernaan Akut Kolera

Profil Penbayar Premi

Jenis Kelamin : Laki-Laki

Pekerjaan : Karyawan

Bidang Pekerjaan : Konstruksi

Gaji : 8 jt.

Rekomendasikan Saya

Reset

Waktu Rekomendasi : 0.39 detik

**Apabila anda memenuhi kondisi pre-existing
Anda diharuskan melakukan medical tes untuk penerimaan asuransi*

Sebaiknya anda memilih asuransi dengan urutan sebagai berikut

1. SPP (Super Protection Plan) Merupakan
2. MHP (Maestro Hospital Plan) merupakan
3. MLP (MaestroLink Plus) Merupakan asura

Gambar 5.6 Rekomendasi Produk SPP

Waktu komputasi yang diperlukan untuk merekomendasikan produk SPP adalah selama 0.39 detik.

Uji coba untuk melakukan rekomendasi produk MLP sebagai pilihan pertama ditunjukkan gambar 5.7

Home Proses Data Mining Rekomendasikan Saya Tentang

Profil Tertanggung

Jenis Kelamin : Laki-Laki

Tanggal Lahir : 2 Feb 1979

Pekerjaan : Karyawan

Bidang Pekerjaan : Jasa/Keuangan

Status Pernikahan : Menikah

Jumlah Anak : 3

Status Merokok : Bukan Perokok

Kondisi Pre-Existing : ☒ Tidak Ada ☐ Ada

**Apabila anda memiliki salah satu penyakit pre-existing dibawah ini, maka pilih button 'Ada'*

Kanker Hepatitis B Hepatitis C Diabetes Mellitus Stroke
Ginjal Pencernaan Akut Kolera

Profil Pembayar Premi

Jenis Kelamin : Laki-Laki

Pekerjaan : Karyawan

Bidang Pekerjaan : Konstruksi

Gaji : 8 jt

Rekomendasikan Saya

Reset

Waktu Rekomendasi : 0.38 detik

**Apabila anda memenuhi kondisi pre-existing
Anda diharuskan melakukan medical tes untuk penerimaan asuransi*

Sebaliknya anda memilih asuransi dengan urutan sebagai berikut

1. MLP
MLP (MaestroLink Plus) Merupakan asuransi
2. MHP
MHP (Maestro Hospital Plan) merupakan
3. SPP
SPP (Super Protection Plan) Merupakan

Gambar 5.7 Rekomendasi Produk MLP

Waktu komputasi yang diperlukan untuk merekomendasikan produk MLP adalah selama 0.38 detik.

Uji coba untuk melakukan rekomendasi produk SK sebagai pilihan pertama ditunjukkan gambar 5.8

Home Proses Data Mining Rekomendasikan Saya Tentang

Profil Tertanggung

Jenis Kelamin : Laki-Laki

Tanggal Lahir : 7 Feb 2013

Pekerjaan : Belum Bekerja

Bidang Pekerjaan : Lainnya

Status Pernikahan : Belum Menikah

Jumlah Anak : 0

Status Merokok : Bukan Perokok

Kondisi Pre-Existing : Tidak Ada

**Apabila anda memiliki salah satu penyakit pre-existing dibawah ini, maka pilih button 'Ada'*

Kanker Hepatitis B Hepatitis C Diabetes Milius Stroke

Singai Pencernaan Akut Kolera

Profil Pembayar Premi

Jenis Kelamin : Laki-Laki

Pekerjaan : Karyawan

Bidang Pekerjaan : Konstruksi

Gaji : 4 jt

Rekomendasikan Saya

Reset

Waktu Rekomendasi : 0.41 detik

**Apabila anda memenuhi kondisi pre-existing*

Anda diharuskan melakukan medical tes untuk penerimaan asuransi

Sebaiknya anda memilih asuransi dengan urutan sebagai berikut

1. SK
2. MHP
3. SPP
4. MLP

SK (Smart Kidz) adalah produk asuransi

MHP (Maestro Hospital Plan) merupakan

SPP (Super Protection Plan) merupakan

MLP (MaestroLink Plus) merupakan asur

Gambar 5.8 Rekomendasi Produk SK

Waktu eksekusi yang diperlukan untuk merekomendasikan produk SK adalah selama 0.41 detik.

Uji coba untuk melakukan rekomendasi produk MHP sebagai pilihan pertama ditunjukkan Gambar 5.9

Home Proses Data Mining Rekomendasikan Saya Tentang

Profil Tertanggung

Jenis Kelamin : Laki-Laki

Tanggal Lahir : 2 Feb 1979

Pekerjaan : Karyawan

Bidang Pekerjaan : Transportasi

Status Pernikahan : Menikah

Jumlah Anak : 3

Status Merokok : Bukan Perokok

Kondisi Pre-Existing : ☒ Tidak Ada ☐ Ada

**Apabila anda memiliki salah satu penyakit pre-existing dibawah ini, maka pilih button 'Ada'*

Kanker Hepatitis B Hepatitis C Diabetes Mellitus Stroke
Ginjal Pencernaan Akut Kolera

Profil Pembayar Premi

Jenis Kelamin : Laki-laki

Pekerjaan : Karyawan

Bidang Pekerjaan : Konstruksi

Gaji : 8 jt

Rekomendasikan Saya

Reset

Waktu Rekomendasi : 0.56 detik

**Apabila anda memenuhi kondisi pre-existing
Anda diharuskan melakukan medical tes untuk penerimaan asuransi*

Sebaliknya anda memilih asuransi dengan urutan sebagai berikut

1. MHP
2. MLP
3. SPP

MHP (Maestro Hospital Plan) merupakan asuransi kesehatan
MLP (MaestroLink Plus) merupakan asuransi kesehatan
SPP (Super Protection Plan) merupakan asuransi kesehatan

Gambar 5.9 Rekomendasi Produk MHP

Waktu komputasi yang diperlukan untuk merekomendasikan produk MHP adalah selama 0.56 detik.

Pada Gambar 5.6, 5.7, 5.9 dapat diketahui bahwa sistem tidak menampilkan produk SK sebagai rekomendasi produk asuransi bagi calon nasabah dikarenakan usia calon nasabah lebih dari 21 tahun dan atau pekerjaannya bukan sorang pelajar/mahasiswa atau belum bekerja, sehingga calon nasabah tersebut tidak memenuhi syarat untuk mengambil produk asuransi SK.

Dari ke-4 pengujian diatas, dapat diketahui bahwa sistem pemilihan jenis asuransi dengan Metode *Naïve Bayes Classifier* mampu merekomendasikan jenis asuransi kepada calon nasabah dengan rata-rata waktu perekomendasi sebesar 0.44 detik.

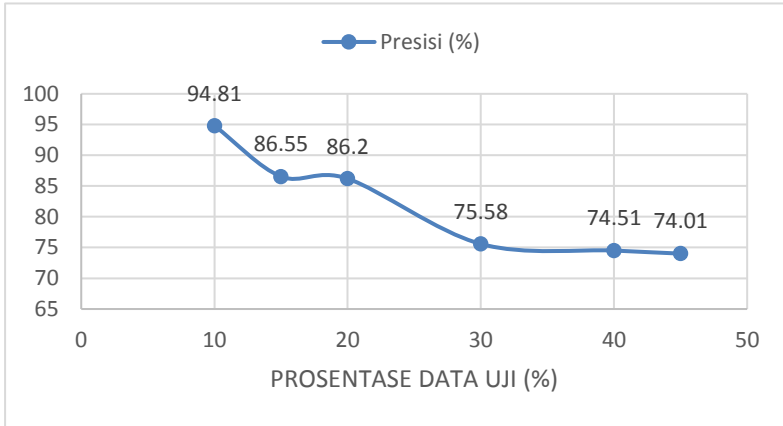
1. Pengujian Tingkat Presisi

Pengujian pertama yang akan diamati merupakan hasil penerapan tingkat presisi. Pengujian tingkat presisi merupakan perbandingan prosentase antara jumlah instance pada data testing yang mampu diklasifikasikan dengan benar dengan keseluruhan jumlah data uji.

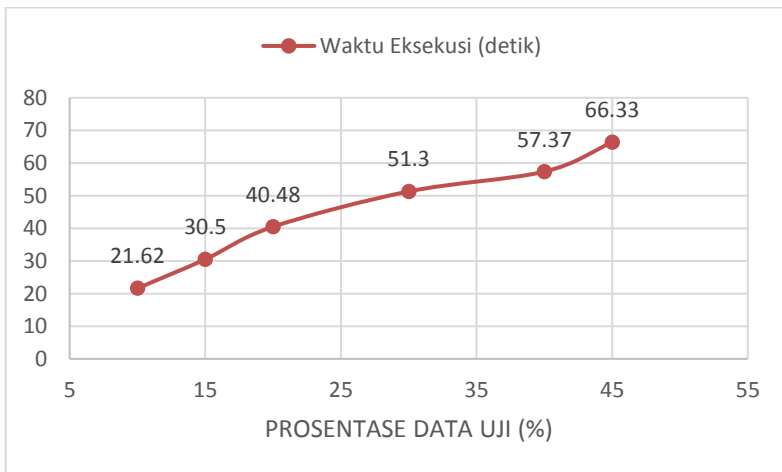
Hasil yang diperoleh dengan melakukan 10-20 kali percobaan dengan perbandingan data latih dan data uji yang berbeda, didapat rata-rata presisi dan waktu eksekusi pengujian pada Tabel 5.2:

Tabel 5.2 Tabel Hasil Presisi Dan Running Time Dengan NBC

Data Latih	Data Uji	Presisi	Waktu Eksekusi
467(90%)	51(10%)	94.81%	21.62 detik
441(85%)	77(15%)	86.55%	30.5 detik
415(80%)	103(20%)	86.2%	40.48 detik
363(70%)	155(30%)	75.58%	51.3 detik
311(60%)	207(40%)	74.51%	57.37 detik
285(55%)	233(45%)	74.01%	66.3 detik



Gambar 5.10 Grafik Prosentase Data Uji Terhadap Rata-Rata Presisi



Gambar 5.11 Grafik Prosentase Data Uji Terhadap Waktu Eksekusi

Gambar 5.10 menunjukkan grafik antara prosentase data uji yang digunakan terhadap hasil rata-rata presisi dari sistem yang telah dibuat. Dari hasil pengujian tersebut dapat disimpulkan bahwa nilai presisi klasifikasi dengan Metode *Naïve Bayes Classifier* akan

cenderung menurun seiring dengan bertambahnya data yang digunakan dalam pengujian. Hasil presisi tertinggi didapatkan ketika menggunakan perbandingan 90% data latih dan 10% data uji, yaitu dengan rata-rata presisi sebesar 94.81% dan rata-rata waktu eksekusi selama 21.62 detik.

Dari Tabel 5.2 dan Gambar 5.11 juga dapat disimpulkan bahwa semakin banyak data yang digunakan sebagai data uji, maka proses pengujian juga akan semakin lama dikarenakan sistem harus melakukan perulangan perhitungan sebanyak jumlah data yang dijadikan data uji serta nilai presisi juga mengalami penurunan seiring bertambahnya proporsi data yang diujikan.

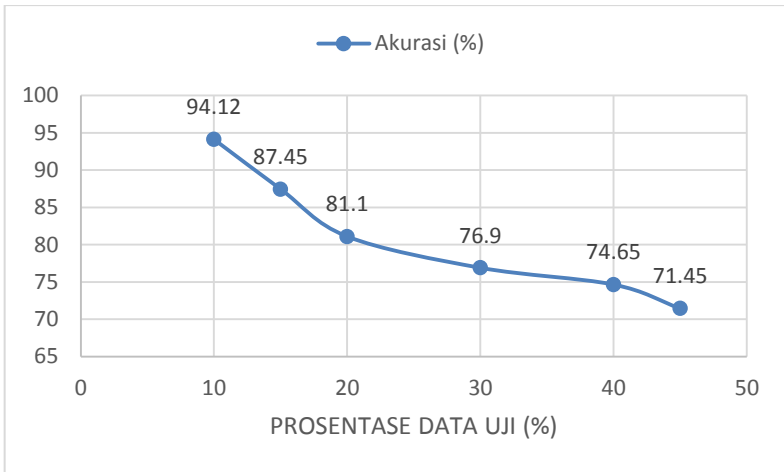
2. Pengujian Tingkat Akurasi

Pengujian kedua yang akan diamati adalah hasil tingkat akurasi. Pengujian presisi merupakan perbandingan antara jumlah prosentasi jumlah instance yang benar setiap kelas klasifikasi dengan jumlah kelas klasifikasi.

Hasil rata-rata akurasi yang didapat setelah melakukan 10-20 kali percobaan dengan proporsi data latih dan data uji yang berbeda. Hasilnya tertera dalam Tabel 5.3 sebagai berikut:

Tabel 5.3 Tabel Hasil Akurasi Klasifikasi Dengan NBC

Data Latih	Data Uji	MLP (%)	SPP (%)	SK (%)	MHP (%)	Akurasi
467(90%)	51(10%)	91	92.9	100	92.58	94.12%
441(85%)	77(15%)	88.9	75.3	100	82.6	87.45%
415(80%)	103(20%)	85.3	53.4	100	85.7	81.1%
363(70%)	155(30%)	84.6	59.38	88.28	67.48	76.9%
311(60%)	207(40%)	86	60.54	97.22	54.86	74.65%
285(55%)	233(45%)	82.72	54.1	98.43	50.59	71.45%



Gambar 5.12 Grafik Prosentase Data Uji Terhadap Rata-Rata Akurasi

Gambar 5.12 menunjukkan grafik antara prosentasi data uji yang digunakan terhadap rata-rata hasil akurasi sistem yang telah dibuat. Dari hasil pengujian tersebut dapat disimpulkan bahwa rata-rata akurasi klasifikasi dengan Metode *Naïve Bayes Classifier* akan cenderung menurun seiring dengan bertambahnya data yang digunakan dalam pengujian. Dari tabel 5.2 dapat juga diketahui bahwa ketika melakukan proses *mining* dengan menggunakan proporsi data latih kurang dari 85%, *Naïve Bayes* tidak cukup baik dalam mengklasifikasikan data uji yang terletak pada kelas SPP. Ini dikarenakan karakteristik demografi nasabah yang memilih produk SPP hampir mirip dengan nasabah yang mengambil produk SK dan kebanyakan nasabah yang mengambil SPP yang masih berusia dibawah 21 tahun dan pekerjaannya masih belum bekerja atau seorang pelajar/mahasiswa akan diklasifikasikan sistem ke dalam kelas SK. Sedangkan rata-rata hasil akurasi tertinggi yaitu 94.12%.

Nilai presisi dan akurasi yang diperoleh memang tidak mampu mencapai 100%. Hal ini dikarenakan seorang nasabah dengan profil demografi yang sama bisa mengambil lebih dari 1 produk

asuransi. Oleh karena itu, dalam implementasi sistem rekomendasi produk asuransi yang dibuat, sistem tidak hanya merekomendasikan satu jenis rekomendasi bagi seorang calon nasabah, melainkan sistem juga akan merekomendasikan semua produk asuransi secara berurutan berdasarkan peringkat probabilitas posterior tiap kelas produk asuransi dan kriteria demografi calon nasabah asuransi harus memenuhi syarat khusus dari produk yang direkomendasikan.

5.3.3 Analisa Hasil Evaluasi

Dari hasil pengujian ke-1 yang telah dilakukan dan berdasarkan Gambar 5.6, Gambar 5.7, Gambar 5.8 dan Gambar 5.9, *Naïve Bayes Classifier* mampu merekomendasikan produk asuransi dengan rata-rata waktu rekomendasi sebesar 0.44 detik. Penerapan Teknik *Laplacian Smoothing* pada proses *training* ketika nilai probabilitas kondisional atributnya bernilai 0 akan menghindarkan munculnya nilai peluang posterior bernilai 0 pada semua kelas klasifikasi yang menyebabkan *classifier* tidak dapat mengklasifikasikan sebuah *record*.

Dari hasil pengujian ke-2 pada Tabel 5.2 dan Tabel 5.3 memperlihatkan hasil eksperimen dengan berbagai proporsi antara data latih dan data uji. Presisi dan akurasi tertinggi *Naïve Bayes Classifier* ketika menggunakan 90% data latih dan 10% data uji dengan rata-rata tingkat presisi sebesar 94.81% dan rata-rata akurasi sebesar 94.12%. Selain itu kinerja sistem akan cenderung menurun seiring dengan berkurangnya data latih, hal ini menunjukkan bahwa semakin besar jumlah data yang digunakan dalam proses *training*, informasi yang diperoleh juga akan semakin banyak sehingga dapat meningkatkan akurasi sistem untuk merekomendasikan data baru yang belum diketahui kelasnya.

Dalam merekomendasikan urutan kelas asuransi pada sebuah *record*, atribut yang paling mempengaruhi urutan jenis asuransi rekomendasi adalah atribut pekerjaan, bidang pekerjaan serta usia dari calon nasabah. Hal ini dikarenakan 3 atribut tersebut memegang peranan penting dalam hal menentukan resiko hidup seseorang.

LAMPIRAN A

Penurunan Formula (2.4)

Diberikan X dan Y variabel acak kontinyu, maka pdf gabungan didefinisikan sebagai berikut:

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-p^2}} \exp \left\{ -\frac{1}{2(1-p^2)} \left[\left(\frac{x-\mu_1}{\sigma_1} \right)^2 - 2p \left(\frac{x-\mu_1}{\sigma_1} \right) \left(\frac{y-\mu_2}{\sigma_2} \right) + \left(\frac{y-\mu_2}{\sigma_2} \right)^2 \right] \right\} \dots\dots\dots (1)$$

$$\begin{aligned} & * \left(\frac{x-\mu_1}{\sigma_1} \right)^2 - 2p \left(\frac{x-\mu_1}{\sigma_1} \right) \left(\frac{y-\mu_2}{\sigma_2} \right) + \left(\frac{y-\mu_2}{\sigma_2} \right)^2 = \left(\frac{x-\mu_1}{\sigma_1} \right)^2 - \\ & 2p \left(\frac{x-\mu_1}{\sigma_1} \right) \left(\frac{y-\mu_2}{\sigma_2} \right) + p^2 \left(\frac{y-\mu_2}{\sigma_2} \right)^2 + (1-p^2) \left(\frac{y-\mu_2}{\sigma_2} \right)^2 \end{aligned}$$

Sehingga persamaan (1) menjadi

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-p^2}} \exp \left\{ -\frac{1}{2(1-p^2)} \left[\left(\frac{x-\mu_1}{\sigma_1} \right)^2 - 2p \left(\frac{x-\mu_1}{\sigma_1} \right) \left(\frac{y-\mu_2}{\sigma_2} \right) + p^2 \left(\frac{y-\mu_2}{\sigma_2} \right)^2 + (1-p^2) \left(\frac{y-\mu_2}{\sigma_2} \right)^2 \right] \right\}$$

*menghitung pdf marginal $f_y(Y)$

$$\begin{aligned} f_y(Y) &= \int_{-\infty}^{\infty} f(x, y) dx \\ &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-p^2}} \exp \left\{ -\frac{1}{2(1-p^2)} \left[\left(\frac{x-\mu_1}{\sigma_1} \right)^2 - 2p \left(\frac{x-\mu_1}{\sigma_1} \right) \left(\frac{y-\mu_2}{\sigma_2} \right) + p^2 \left(\frac{y-\mu_2}{\sigma_2} \right)^2 + (1-p^2) \left(\frac{y-\mu_2}{\sigma_2} \right)^2 \right] \right\} dx \end{aligned}$$

LAMPIRAN A (LANJUTAN)

$$= \int_{-\infty}^{\infty} \frac{\exp\left\{-\frac{(1-p^2)\left(\frac{y-\mu_2}{\sigma_2}\right)^2}{2(1-p^2)}\right\}}{2\pi\sigma_1\sigma_2\sqrt{1-p^2}} \exp\left\{-\frac{1}{2(1-p^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right) + p\left(\frac{y-\mu_2}{\sigma_2}\right)\right]^2\right\} dx$$

$$P(X|Y) = \int_{-\infty}^{\infty} \frac{\exp\left\{-\frac{\left(\frac{y-\mu_2}{\sigma_2}\right)^2}{2}\right\}}{2\pi\sigma_1\sigma_2\sqrt{1-p^2}} \exp\left\{-\frac{1}{2(1-p^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right) + p\left(\frac{y-\mu_2}{\sigma_2}\right)\right]^2\right\} dx$$

Misalkan

$$a = \frac{x-\mu_1}{\sigma_1}$$

$$b = \frac{y-\mu_2}{\sigma_2}$$

$$da = \frac{dx}{\sigma_1}$$

batas:

$$\sigma_1 da = dx$$

$$x \rightarrow \infty \quad a \rightarrow \infty$$

$$x \rightarrow -\infty \quad a \rightarrow -\infty$$

Sehingga $f_y(Y)$ dapat dituliskan kembali menjadi:

$$f_y(Y) = \frac{\sigma_2 \exp\left\{-\frac{1}{2}b^2\right\}}{\sigma_1\sigma_2\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(1-p^2)}} \exp\left\{-\frac{1}{2(1-p^2)}(a-pb)^2\right\} da$$

Karena $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(1-p^2)}} \exp\left\{-\frac{1}{2(1-p^2)}(a-pb)^2\right\} da$ merupakan pdf dari $N(pb, (1-p^2))$ jadi nilainya adalah 1.

$$f_y(Y) = \frac{\exp\left\{-\frac{1}{2}b^2\right\}}{\sigma_1\sqrt{2\pi}} \dots\dots\dots(2)$$

LAMPIRAN A (LANJUTAN)

Dihitung probabilitas kondisional $P(X|Y)$

$$P(X|Y) = \frac{f(x,y)}{f_y(Y)}$$

$$P(X|Y) = \frac{\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-p^2}} \exp\left\{-\frac{1}{2(1-p^2)}[a^2 - 2pab + b^2]\right\}}{\frac{\exp\left\{-\frac{1}{2}b^2\right\}}{\sigma_1\sqrt{2\pi}}}$$

$$P(X|Y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-p^2}} \exp\left\{-\frac{1}{2(1-p^2)}[a^2 - 2pab + b^2]\right\} (2\sigma_2\sqrt{2\pi}) \left(\exp\left\{\left(\frac{1}{2}b^2\right)\right\}\right)$$

$$P(X|Y) = \frac{1}{\sigma_1\sqrt{2\pi(1-p^2)}} \exp\left\{-\frac{1}{2(1-p^2)}[a^2 - 2pab + b^2 + (p^2 - 1)b^2]\right\}$$

$$P(X|Y) = \frac{1}{\sigma_1\sqrt{2\pi(1-p^2)}} \exp\left\{-\frac{1}{2(1-p^2)}[a^2 - 2pab + p^2b^2]\right\}$$

$$P(X|Y) = \frac{1}{\sigma_1\sqrt{2\pi(1-p^2)}} \exp\left\{-\frac{1}{2(1-p^2)}[a - pb]^2\right\}$$

$$P(X|Y) = \frac{1}{\sigma_1\sqrt{2\pi(1-p^2)}} \exp\left\{-\frac{1}{2(1-p^2)}[a - pb]^2\right\}$$

$$P(X|Y) = \frac{1}{\sigma_1\sqrt{2\pi(1-p^2)}} \exp\left\{-\frac{1}{2(1-p^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right) - p\left(\frac{y-\mu_2}{\sigma_2}\right)\right]^2\right\}$$

$$P(X|Y) = \frac{1}{\sigma_1\sqrt{2\pi(1-p^2)}} \exp\left\{-\frac{1}{2(1-p^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right) - p\left(\frac{y-\mu_2}{\sigma_2}\right)\right]^2\right\}$$

$$P(X|Y) = \frac{1}{\sigma_1\sqrt{2\pi(1-p^2)}} \exp\left\{-\frac{1}{2(1-p^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right) - p\left(\frac{y-\mu_2}{\sigma_2}\right)\right]^2\right\}$$

LAMPIRAN A (LANJUTAN)

$$P(X|Y) = \frac{1}{\sigma_1 \sqrt{2\pi(1-p^2)}} \exp \left\{ -\frac{1}{2(1-p^2)} \left[\left(\frac{1}{\sigma_1} \right) (x - \mu_1) - p\sigma_1 \left(\frac{y - \mu_2}{\sigma_2} \right) \right]^2 \right\}$$

$$P(X|Y) = \frac{1}{\sigma_1 \sqrt{2\pi(1-p^2)}} \exp \left\{ -\frac{1}{2\sigma_1^2(1-p^2)} \left[(x - \mu_1) - p\sigma_1 \left(\frac{y - \mu_2}{\sigma_2} \right) \right]^2 \right\}$$

$$P(X|Y) = \frac{1}{\sqrt{2\pi\sigma_1^2(1-p^2)}} \exp \left\{ -\frac{1}{2\sigma_1^2(1-p^2)} \left[x - \left(\mu_1 + p \frac{\sigma_1}{\sigma_2} (y - \mu_2) \right) \right]^2 \right\} \dots\dots\dots (3)$$

$$P(X|Y) \sim N \left(\mu_1 + p \frac{\sigma_1}{\sigma_2} (y - \mu_2), \sigma_1^2(1 - p^2) \right)$$

- Karena X dan Y independen maka $p = 0$, sehingga persamaan (3) menjadi:

$$P(X|Y) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp \left\{ -\frac{1}{2\sigma_1^2} (x - \mu_1)^2 \right\}$$

Dengan kata lain $P(X|Y)$ berdistribusi normal dengan $\mu = \mu_1$ dan $\sigma = \sigma_1$

$$P(X|Y) \sim N(\mu_1, \sigma_1^2)$$

LAMPIRAN B

Output Sistem Pemilihan Jenis Asuransi

- a. Halaman Utama (Home)



Pada menu *Home*, terdapat tampilan utama sistem perekomendasi ini. Didalamnya juga ada beberapa menu yang bisa diakses *user*, yaitu menu ‘Data Historis’ untuk menampilkan keseluruhan data historis nasabah asuransi serta mengupdate isi data tersebut. Pada halaman ‘Data Atribut Nasabah’, *user* dapat menambah atau mengurangi kategori setiap atribut yang dimiliki oleh nasabah asuransi, selain itu *user* juga dapat merubah batas minimum ‘gaji’ dari pembayar premi asuransi. Sedangkan dalam menu ‘Proses Data Mining’, *user* mampu melakukan semua fungsi yang sudah dijelaskan di buku ini.

Pada halaman ini *user* bisa membaca data mentah polis nasabah asuransi dan melakukan perubahan terhadap isi datanya serta melakukan penyisipan data baru.






Insert New Data

No Polis :
 Jenis Kelamin :

P

 Tanggal Lahir :

0

Usia Masuk :

0

 Tahun
 Pekerjaan :

Bumh Beryaga

 Bidang Pekerjaan :

Lainnya

Gaji Pemegang Rekening :

0

 Rp
 Status Pernikahan :

Menikah

 Jumlah Anak :

0

 Status Perkawinan :

Perokok

 Riwayat Kesehatan :

Tidak Ada

Plan Asuransi :

MLP

LAMPIRAN B (LANJUTAN)

c. Halaman Proses *Data Mining*

Seluruh operasi yang terdapat dalam halaman ini sudah dijelaskan secara rinci di buku ini dalam bab 4 dan bab 5.

d. Halaman Data Atribut

Tabel Kategori Tiap Atribut

Atribut	Kategori
children	Banyak
children	Normal
children	Tidak Ada
field	Pemerintahan
field	Transportasi
field	Perdagangan
field	Konstruksi
field	Manufaktur
field	SDA
field	Jasa/Keuangan
field	Lainnya
gender	L
gender	P
job	TNI/Polri
job	Ibu Rumah Tangga
job	Karyawan
job	Mahasiswa/Pelajar
job	Pemilik Usaha

Atribut :

Kategori :

Set Minimum Salary

Rp

Pada halaman ini akan ditampilkan masing-masing kategori dari tiap atribut nasabah asuransi, dimana penambahan kategori pada tiap atribut juga dapat

LAMPIRAN B (LANJUTAN)

dilakukan pada halaman ini. Selain itu, *user* juga dapat mengatur batas minimum gaji/pendapatan pembayar premi yang diperbolehkan mengikuti program asuransi.

LAMPIRAN C

Pengujian dengan Merubah Nilai Atribut

- a. Perubahan Nilai Gaji Pembayar Premi
1. Diberikan satu *record* data calon nasabah asuransi sebagai berikut.

Jenis Kelamin	: Laki-laki
Usia	: 9 tahun
Pekerjaan	: Mahasiswa/Pelajar
Bidang Pekerjaan	: Lainnya
Gaji Pembayar Premi	: 3 jt/per bulan
Status Pernikahan	: Belum Menikah
Jumlah Anak	: 0
Status Merokok	: Bukan Perokok
Kondisi <i>Pre-Existing</i>	: Tidak Ada

Hasil Urutan Rekomendasi Produk Asuransi

Home Proses Data Mining Rekomendasikan Saya Tentang

Profil Tertanggung

Jenis Kelamin : Laki-Laki

Tanggal Lahir : 6 Jul 2007

Pekerjaan : Mahasiswa/Pelajar

Bidang Pekerjaan : Lainnya

Status Pernikahan : Belum Menikah

Jumlah Anak : 0

Status Merokok : Bukan Perokok

Kondisi Pre-Existing : ☒ Tidak Ada ☐ Ada

**Apabila anda memiliki salah satu penyakit pre-existing di bawah ini, maka pilih button 'Ada'*

Kanker Hepatitis B Hepatitis C Diabetes Mellitus Stroke Gigitan Pencernaan Akut Kolera

Profil Pembayar Premi

Jenis Kelamin : Perempuan

Pekerjaan : Karyawan

Bidang Pekerjaan : Jasa/Keuangan

Gaji : 3 jt

Rekomendasikan Saya

Reset

Waktu Rekomendasi : 0.77 detik

**Apabila anda memunculkan kondisi pre-existing, Anda disarankan melakukan medical tes untuk pemeriksaan asuransi*

Sebaiknya anda memilih asuransi dengan urutan sebagai berikut:

1. SK
2. SPP
3. MHP
4. MLP

SK (Smart Kids) adalah produk asuransi
SPP (Super Protection Plan) merupakan
MHP (Maestro Hospital Plan) merupakan
MLP (MaestroLink Plus) merupakan asuransi

Urutan rekomendasi produk adalah 1. SK, 2. SPP, 3. MHP, dan yang terakhir adalah 4. MLP.

LAMPIRAN C (LANJUTAN)

2. Diberikan *record* yang sama dengan kasus 1, hanya saja pada atribut gaji pembayar premi, jumlahnya diganti 20 juta/bulan

Hasil Urutan Rekomendasi Produk Asuransi

Profil Tertanggung

Jenis Kelamin : Laki-Laki

Tanggal Lahir : 6 Juli 2007

Pekerjaan : Mahasiswa/Pelajar

Bidang Pekerjaan : Lainnya

Status Pernikahan : Belum Menikah

Jumlah Anak : 0

Status Merokok : Bukan Perokok

Kondisi Pre-Existing : Tidak Ada

*Apabila anda memiliki salah satu penyakit pre-existing dibawah ini, maka pilih button 'Ada'

Kanker Hepatitis B Hepatitis C Diabetes Mellitus Stroke
Ginjal Pencernaan Akut Kolera

Profil Pembayar Premi

Jenis Kelamin : Perempuan

Pekerjaan : Karyawan

Bidang Pekerjaan : Jasa/Keuangan

Gaji : 20

Rekomendasikan Saya

Reset

Waktu Rekomendasi : 0.45 detik

*Apabila anda memenuhi kondisi pre-existing
Anda diharuskan melakukan medical tes untuk penerimaan asuransi

Setidaknya anda memilih asuransi dengan urutan sebagai berikut

1. SPP
2. SK
3. MHP
4. MLP

SPP (Super Protection Plan) merupakan SK (Smart Kids) adalah produk asuransi MHP (Maestro Hospital Plan) merupakan MLP (Maestro Clinic Plus) merupakan asuransi

Urutan rekomendasi produk adalah 1. SPP, 2. SK, 3. MHP, dan yang terakhir adalah 4. MLP.

LAMPIRAN C (LANJUTAN)

3. Diberikan *record* yang sama dengan kasus 1, hanya saja pada atribut gaji pembayar premi, jumlahnya diganti 30 juta/bulan

Hasil Urutan Rekomendasi Produk Asuransi

Home Proses Data Mining Rekomendasikan Saya Tentang

Profil Tertanggung

Jenis Kelamin : Laki-Laki

Tanggal Lahir : 6 Jul 2007

Pekerjaan : Mahasiswa/Pelajar

Bidang Pekerjaan : Lainnya

Status Pernikahan : Belum Menikah

Jumlah Anak : 0

Status Merokok : Bukan Perokok

Kondisi Pre-Existing : Tidak Ada

**Apabila anda memiliki salah satu penyakit pre-existing dibawah ini, maka pilih button 'Ada'*

Kanker Hepatitis B Hepatitis C Diabetes Melitus Stroke
Ginjal Pencernaan Akut Kolera

Profil Pembayar Premi

Jenis Kelamin : Perempuan

Pekerjaan : Karyawan

Bidang Pekerjaan : Jasa/Keuangan

Gaji : 30

Rekomendasikan Saya

Reset

Waktu Rekomendasi : 0.72 detik

**Apabila anda memiliki kondisi pre-existing
Anda diharuskan melakukan medical tes untuk penerimaan asuransi*

Setidaknya anda memilih asuransi dengan urutan sebagai berikut

1. SPP
2. MHP
3. MLP
4. SK

SK (Smart Kidz) adalah produk asuransi

Urutan rekomendasi produk adalah 1. SPP, 2. MHP, 3. MLP, dan yang terakhir adalah 4. SK.

LAMPIRAN C (LANJUTAN)

- b. Perubahan Kategori Pada Pekerjaan, Bidang Pekerjaan dan Kondisi *pre-existing*
4. Diberikan satu *record* data calon nasabah asuransi sebagai berikut.

Jenis Kelamin	: Laki-laki
Usia	: 35 tahun
Pekerjaan	: Karyawan
Bidang Pekerjaan	: Jasa/Keuangan
Gaji Pembayar Premi	: 5 jt/per bulan
Status Pernikahan	: Menikah
Jumlah Anak	: 2
Status Merokok	: Bukan Perokok
Kondisi <i>Pre-Existing</i>	: Tidak Ada

Hasil Urutan Rekomendasi Produk Asuransi

The screenshot shows a web application with a navigation bar at the top containing 'Home', 'Proses Data Mining', 'Rekomendasikan Saya', and 'Tentang'. The main content area is divided into two columns. The left column, titled 'Profil Tertanggung', contains a form with the following fields and values: Jenis Kelamin (Laki-Laki), Tanggal Lahir (6 Jul 1981), Pekerjaan (Karyawan), Bidang Pekerjaan (Jasa/Keuangan), Status Pernikahan (Menikah), Jumlah Anak (2), Status Merokok (Bukan Perokok), and Kondisi Pre-Existing (Tidak Ada). The right column, titled 'Profil Pembayar Premi', contains a form with: Jenis Kelamin (Perempuan), Pekerjaan (Karyawan), Bidang Pekerjaan (Jasa/Keuangan), and Gaji (5 jt). Below these forms are buttons for 'Rekomendasikan Saya' and 'Reset'. To the right of the 'Profil Pembayar Premi' form, a text box lists recommended insurance products: 'Sebaiknya anda memilih asuransi dengan urutan sebagai berikut: 1. MLP (MaestroLink Plus) Merupakan asuransi, 2. SPP (Super Protection Plan) Merupakan, 3. MHP (Maestro Hospital Plan) merupakan'. At the bottom, a status bar shows 'Waktu Rekomendasi : 0.50 detik' and a note: '*Apabila anda memiliki kondisi pre-existing' and 'Anda disarankan melakukan medical tes untuk penerimaan asuransi.'

Urutan rekomendasi produk adalah 1. MLP, 2. SPP, dan yang terakhir adalah 3. MHP.

LAMPIRAN C (LANJUTAN)

5. Diberikan *record* yang sama dengan kasus 1, hanya saja pada atribut bidang pekerjaan, kategori diubah menjadi konstruksi.

Hasil Urutan Rekomendasi Produk Asuransi

Home Proses Data Mining Rekomendasi Saya Tentang

Profil Tertanggung

Jenis Kelamin : Laki-Laki

Tanggal Lahir : 6 Jul 1981

Pekerjaan : Karyawan

Bidang Pekerjaan : Konstruksi

Status Pernikahan : Menikah

Jumlah Anak : 2

Status Merokok : Bukan Perokok

Kondisi Pre-Existing : Tidak Ada

**Apabila anda memiliki salah satu penyakit pre-existing dibawah ini, maka pilih button 'Ada'*

Kanker Hepatitis B Hepatitis C Diabetes Mellitus Stroke

Gigitan Pencernaan Akut Kolera

Profil Pembayar Premi

Jenis Kelamin : Perempuan

Pekerjaan : Karyawan

Bidang Pekerjaan : Jasa Keuangan

Gaji : 5

Rekomendasikan Saya

Reset

Waktu Rekomendasi : 0.55 detik

Apabila anda memenuhi kondisi pre-existing, Anda diharuskan melakukan medical tes untuk penerimaan asuransi

Selanjutnya anda memilih asuransi dengan urutan sebagai berikut

1. SPP (Super Protection Plan) Merupakan
2. MLP (MaestroLink Plus) Merupakan asur
3. MHP (Maestro Hospital Plan) merupaka

Urutan rekomendasi produk adalah 1. SPP, 2. MLP, S dan yang terakhir adalah 3. MHP.

LAMPIRAN C (LANJUTAN)

6. Diberikan *record* yang sama dengan kasus 1, hanya saja pada atribut pekerjaan diganti dengan pemilik usaha, bidang pekerjaan kategorinya diubah menjadi Perdagangan dan ada kondisi *pre-existing*.

Hasil Urutan Rekomendasi Produk Asuransi

The screenshot shows a web application for insurance recommendation. The interface is divided into two main sections: 'Profil Tertanggung' (Insured Profile) and 'Profil Pembayar Premi' (Premium Payer Profile). The 'Profil Tertanggung' section includes fields for Gender (Male), Birth Date (5 Jul 1981), Occupation (Owner/Entrepreneur), Field of Occupation (Trade), Marital Status (Married), Number of Children (2), Smoking Status (Non-Smoker), and Pre-existing Conditions (None). The 'Profil Pembayar Premi' section includes fields for Gender (Female), Occupation (Employee), Field of Occupation (Finance), and Salary (5). A 'Rekomendasikan Saya' button is located between the two sections. Below the buttons, there is a 'Waktu Rekomendasi' (Recommendation Time) field showing 0.42 detik. A list of recommended insurance products is shown on the right: 1. MHP (Maestro Hospital Plan), 2. MLP (MaestroLink Plus), and 3. SPP (Super Protection Plan). A note at the bottom states: 'Anda diharuskan melakukan medical tes untuk penerimaan asuransi' (You are required to undergo medical tests for insurance acceptance).

Urutan rekomendasi produk adalah 1. MHP, 2. MLP, dan yang terakhir adalah 3. SPP.

LAMPIRAN D

Data Polis Nasabah Asuransi

Source: AXA Financial Indonesia Kantor Cabang Gresik mulai bulan Januari 2014- Januari2015.

Jumlah Polis Nasabah Asuransi: 518 *instances*

Keterangan Atribut:

1. X1= Jenis Kelamin Tertanggung (Tipe: Kategorik)
2. X2= Usia Masuk Tertanggung (Tipe: Numerik)
3. X3= Pekerjaan Tertanggung (Tipe: Kategorik)
4. X4= Bidang Pekerjaan Tertanggung (Tipe: Kategorik)
5. X5= Rata-rata Gaji Pembayar Premi per bulan dalam satuan juta (Tipe: Numerik)
6. X6= Status Pernikahan Tetanggung (Tipe: Kategorik)
7. X7= Jumlah Anak Tertanggung (Tipe: Kategorik)
8. X8= Status Merokok Tertanggung (Tipe: Kategorik)
9. X9= Kondisi ada tidaknya penyakit *Pre-existing* Tertanggung (Tipe: Kategorik)
10. Y= Jenis Asuransi.

Pada halaman selanjutnya, akan dilampirkan sebagian data yang digunakan dalam penelitian Tugas Akhir ini.

LAMPIRAN D (LANJUTAN)

No. Polis	X1	X2	X3	X4	X5	X6	X7	X8	X9	Y
508-7052519	L	2	Belum Bekerja	Lainnya	10	Belum Menikah	0	T	Tidak Ada	SK
508-7054739	P	17	Mahasiswa Pelajar	Lainnya	40	Belum Menikah	0	T	Tidak Ada	MLP
508-6976288	P	13	Mahasiswa Pelajar	Lainnya	40	Belum Menikah	0	T	Tidak Ada	MLP
508-6976254	L	20	Mahasiswa Pelajar	Lainnya	40	Belum Menikah	0	T	Tidak Ada	MLP
508-2727628	L	55	Karyawan	Perdagangan	10	Menikah	3	T	Tidak Ada	MLP
508-1219510	L	35	TNI/Polri	Pemerintahan	60	Menikah	2	Y	Tidak Ada	SPP
508-6976296	P	49	Pemilik Usaha	Perdagangan	6	Menikah	2	T	Tidak Ada	SPP
508-2256164	L	46	Karyawan	Transportasi	12	Menikah	2	T	Tidak Ada	MLP
508-7936760	L	42	Karyawan	Manufaktur	7	Menikah	0	Y	Tidak Ada	SPP
508-7052568	L	29	Pemilik Usaha	Jasa/keuangan	7	Belum Menikah	0	Y	Tidak Ada	MLP
508-6976338	L	8	Mahasiswa Pelajar	Lainnya	10	Belum Menikah	0	T	Tidak Ada	SPP
508-4996452	L	40	Pemilik Usaha	Perdagangan	7	Menikah	3	Y	Tidak Ada	MHP

LAMPIRAN D (LANJUTAN)

508-7172598	P	29	Pemilik Usaha	Jasa Keuangan	15	Menikah	1	T	Ada	MHP
508-7936877	P	4	Belum Bekerja	Lainnya	7	Belum Menikah	0	T	Tidak Ada	SPP
508-7936885	L	36	Karyawan	Jasa Keuangan	7	Menikah	1	Y	Tidak Ada	MLP
508-8059539	P	45	Karyawan	Jasa Keuangan	7	Menikah	2		Tidak Ada	MHP
508-7936760	L	42	Karyawan	Manufaktur	7	Belum Menikah	0	Y	Tidak Ada	SPP
508-7052618	L	46	Karyawan	Jasa Keuangan	20	Menikah	3	T	Tidak Ada	MLP
508-7221841	P	41	Pemilik Usaha	Perdagangan	20	Menikah	3		Ada	MHP
508-6976338	P	36	Pemilik Usaha	Perdagangan	10	Menikah	2	T	Tidak Ada	SPP
508-7172598	P	29	Pemilik Usaha	Jasa Keuangan	15	Menikah	1		Tidak Ada	MHP
508-6975462	L	3	Belum Bekerja	Lainnya	3	Belum Menikah	0	T	Tidak Ada	MLP
508-7936778	P	13	Mahasiswa Pelajar	Lainnya	7	Belum Menikah	0	T	Tidak Ada	SPP
508-7936885	L	36	Karyawan	Jasa Keuangan	7	Menikah	1	Y	Tidak Ada	MLP
508-4996577	L	14	Mahasiswa Pelajar	Lainnya	6	Belum Menikah	0	T	Ada	MHP

LAMPIRAN D (LANJUTAN)

508-7936919	P	37	Karyawan	Jasa/keuangan	7	Menikah	3	T	Tidak Ada	MLP
508-3130558	P	28	Pemilik Usaha	Perdagangan	6	Menikah	3	T	Ada	MLP
508-7936901	P	38	Karyawan	Jasa/keuangan	7	Menikah	1	T	Tidak Ada	MLP
508-8612105	L	13	Mahasiswa/Pelajar	Lainnya	5	Belum Menikah	0	T	Tidak Ada	MLP
508-6977435	L	20	Karyawan	Konstruksi	4	Belum Menikah	0	Y	Tidak Ada	MLP
508-8612055	L	32	Karyawan	Manufaktur	6	Belum Menikah	0	T	Tidak Ada	MLP
508-4504819	P	10	Mahasiswa/Pelajar	Lainnya	5	Belum Menikah	0	T	Tidak Ada	SPP
508-8612022	P	4	Belum Bekerja	Lainnya	10	Belum Menikah	0	T	Tidak Ada	SPP
508-7936927	L	31	Karyawan	Perdagangan	6	Belum Menikah	0	T	Tidak Ada	MLP
508-8612030	P	37	Karyawan	Pemerintahan	5	Menikah	1	T	Tidak Ada	MLP
508-7309795	P	30	Karyawan	Perdagangan	7	Menikah	0		Tidak Ada	MHP
508-7936950	P	24	Karyawan	Manufaktur	6	Belum Menikah	0	T	Tidak Ada	MLP
508-7936968	L	10	Mahasiswa/Pelajar	Lainnya	10	Belum Menikah	0	T	Tidak Ada	MLP

LAMPIRAN D (LANJUTAN)

508-8612477	L	20	Pemilik Usaha	Jasa keuangan	5	Belum Menikah	0	Y	Tidak Ada	NLP
508-8633994	L	45	Karyawan	SDA	5	Menikah	2	Y	Ada	NLP
508-7105721	P	1	Belum Bekerja	Lainnya	5	Belum Menikah	0	T	Tidak Ada	SK
508-8612048	P	2	Belum Bekerja	Lainnya	5	Belum Menikah	0	T	Tidak Ada	SK
508-7105705	P	7	Mahasiswa Pelajar	Lainnya	6	Belum Menikah	0	T	Tidak Ada	SPP
508-8612477	L	20	Pemilik Usaha	Jasa keuangan	5	Belum Menikah	0	Y	Tidak Ada	NLP
508-8010987	P	47	Karyawan	Jasa keuangan	5	Menikah		T	Ada	MHP
508-7105655	P	26	Karyawan	Pemerintahan	7	Belum Menikah	0	T	Tidak Ada	NLP
508-4857290	L	56	Karyawan	Jasa keuangan	5	Menikah	3	Y	Ada	NLP
508-4504835	L	6	Mahasiswa Pelajar	Lainnya	5	Belum Menikah	0	T	Tidak Ada	NLP
508-8634018	P	5	Belum Bekerja	Lainnya	5	Belum Menikah	0	T	Tidak Ada	SPP
508-8634026	L	7	Mahasiswa Pelajar	Lainnya	5	Belum Menikah	0	T	Tidak Ada	SPP
508-6977443	P	52	Karyawan	Perdagangan	5	Menikah	3	T	Tidak Ada	NLP

508-7105689	L	39	Karyawan	Manufaktur	7	Menikah	2	Y	Tidak Ada	MLP
508-8060032	L	47	Karyawan	Jasa keuangan	10	Duda/Janda	2	T	Tidak Ada	MHP
508-8612469	L	7	Mahasiswa Pelajar	Lainnya	5	Belum Menikah	0	T	Tidak Ada	MLP
508-8634000	P	25	Karyawan	Jasa keuangan	6	Belum Menikah	0	T	Tidak ada	MLP
508-6977427	P	49	Karyawan	Konstruksi	15	Menikah	2	T	Tidak Ada	MLP
508-8060255	P	28	Pemilik Usaha	Perdagangan	6	Menikah	2		Tidak Ada	MHP
308-7488387	P	8	Mahasiswa Pelajar	Lainnya	8	Belum Menikah	0	T	Tidak Ada	MLP
508-8612089	P	4	Belum Bekerja	Lainnya	10	Belum Menikah	0	T	Tidak Ada	MLP
508-7797238	L	1	Belum Bekerja	Lainnya	5	Belum Menikah	0		Tidak Ada	MHP
508-8634026	L	7	Mahasiswa Pelajar	Lainnya	5	Belum Menikah	0	T	Tidak Ada	SPP
508-8641237	P	8	Mahasiswa Pelajar	Lainnya	6	Belum Menikah	0	T	Tidak Ada	SPP
508-7797279	L	6	Mahasiswa Pelajar	Lainnya	5	Belum Menikah	0		Tidak Ada	MHP
508-8634208	L	46	Karyawan	Konstruksi	5	Menikah	2	T	Tidak Ada	MLP

LAMPIRAN D (LANJUTAN)

LAMPIRAN D (LANJUTAN)

LAMPIRAN D (LANJUTAN)

BAB VI PENUTUP

Pada bab ini diuraikan beberapa hal yang dapat disimpulkan dari penyelesaian Tugas Akhir ini sekaligus beberapa kemungkinan pengembangannya

6.1 Kesimpulan

Dari hasil uji coba sistem dapat diambil beberapa kesimpulan sebagai berikut:

1. Perangkat lunak yang dihasilkan mampu melakukan *preprocessing*, *data mining* dan merekomendasikan jenis asuransi bagi calon nasabah baru dengan mengimplementasikan Metode *Naïve Bayes Classifier*.
2. Metode *Naïve Bayes Classifier* mampu mengklasifikasikan nasabah asuransi berdasarkan ciri demografinya dengan rata-rata akurasi tertinggi sebesar 94.12% ketika pembagian data latih 90% dan data uji 10% dari total *dataset* sebanyak 518 data polis nasabah asuransi.

6.2 Saran

Dalam penelitian selanjutnya sebaiknya setiap atribut profil nasabah asuransi diberikan bobot dan perangkat lunak yang dikembangkan selanjutnya bisa menentukan uang premi dan uang pertanggungan yang direkomendasikan kepada calon nasabah asuransi.

DAFTAR PUSTAKA

- [1] Abdulkadir, M., *Hukum Asuransi Indonesia*. PT Citra Aditya Bakti. Bandung (2006).
- [2] Bakar, A. A., Othman, Z., Yusoff M. S. N. M., & Ismail, R, "Development of Knowledge Model for Insurance Product Decision using the Associative Classification Approach,," In 10th International on Intelligent System Design and Application, Malaysia (2010).
- [3] Xing, B., Jian, L., & Feng-wen, H, "The Application of Improved BP Algorithm in Customer Classification of Life Insurance,," In 16th International Conference on Management Science & Engineering. China (2009).
- [4] Karatabak, M., "A New Classifier for Breast Cancer Detection Based on Naïve Bayes Classifier,," *Sciencedirect*. Vol. 72. (2015, Mar) 32-36.
- [5] Kusrini & Luthfi, E. T., *Algoritma Data Mining*. Andi Offset. Yogyakarta (2009).
- [6] Kusumadewi, S, "Klasifikasi Status Gizi Menggunakan Naïve Bayes Classification,," in *Communication and Information Technology Journal*, Vol: 3 (2009) 6-11.
- [7] Han, J., & Kamber, M, *Data Mining: Concepts and Techniques*, 3rd Edition. Morgan Kauffman Publisher, San Francisco (2012).
- [8] Bustami, B. *Penerapan Algoritma Naïve Bayes Untuk Mengklasifikasi Data Nasabah Asuransi*. TECHSI. Jurnal Penelitian Teknik Informatika, Vol: 3 No. 2 (2013) 127-146.
- [9] Mariana, dkk. *Analisis Penanganan Missing Value dengan Metode CMVE*. Universitas Telkom (2011).

BIODATA PENULIS



Penulis memiliki nama lengkap Lailatul Mabadi Chaira dengan nama panggilan Ella, lahir di kota Gresik pada tanggal 4 November 1994. Penulis berasal dari Kota Gresik, bertempat tinggal di Jalan Satelit 08 No. 07 Manyar Gresik. Pendidikan formal yang pernah ditempuh yaitu MI Banat Manyar Gresik, SMP Negeri 1 Manyar, dan SMA Negeri 1 Gresik. Kemudian, penulis melanjutkan studi di jurusan Matematika ITS, dengan bidang minat ilmu komputer. Dalam bidang minat ini penulis mulai mengenal bahasa pemrograman diantaranya adalah C, C++, Java, PHP-MySQL, dan MATLAB. Semasa menempuh jenjang pendidikan S-1, penulis juga aktif dalam kegiatan non-akademis diantaranya aktif di organisasi kemahasiswaan Matematika ITS. Selama penulisan tugas akhir ini Penulis tidak lepas dari kekurangan, untuk itu penulis mengharapkan kritik, saran, dan pertanyaan mengenai Tugas Akhir ini yang dapat dikirimkan melalui *e-mail* ke ellachaira@gmail.com.